

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:26:23 ; Search time 3097 Seconds
(without alignments)
1954.997 Million cell updates/sec

Title: US-09-980-054A-12

Perfect score: 772

Sequence: 1 MSDIDIDNVLNLEEQYELG.....NLVKEVGGLQVSENPDMM 148

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cpn2.1/USPTO.spool/US09980054/runat.13022004.163839.26298/app.query.fasta_1.327
-DB=GenEmbl -OFT=fastap -SUFFIX=rs -MINMATCH=0 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=ext -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054@CGN.1.1.3508.0/runat.13022004.163839.26298 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.ma.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*

RESULT 1

29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.man.*
37: em.htg.vrt.*
38: em.sy.*
39: em.higo.hum.*
40: em.higo.mus.*
41: em.higo.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	772	100.0	447	6	AX057345	AX057345 Sequence
2	764	99.0	447	6	AX240813	AX240813 Sequence
3	764	99.0	447	6	AX488762	AX488762 Sequence
4	562	72.8	326	6	AX057353	AX057353 Sequence
5	223	28.9	1104	8	SCYNL260C	271536 S.cerevisia
6	223	28.9	2961	8	YSCATX1HF	L35270 Saccharomyc
7	223	28.9	33016	8	SCCHXIVL	X96722 S.cerevisia
8	132	17.1	25000	8	SPCC191	AL048644 S.pombe c
9	130	16.8	105692	2	AP003953	AP003953 Oryza sat
10	130	16.8	144741	8	AP004674	AP004674 Oryza sat
11	122.5	15.9	96283	8	NC20H10	BX294024 Neurospor
12	111	14.4	1904	8	AF360255	AF360255 Arabidops
13	111	14.4	129667	8	AC006234	AC006234 Arabidops
14	108	14.0	58350	2	AC137683	AC137683 Homo sapi
15	107.5	13.9	30851	1	AE001581	AE001581 Borrelia
16	107	13.9	65	6	AX485584	AX485584 Sequence
17	104	13.5	29838	1	AE001578	AE001578 Borrelia
18	101.5	13.1	257109	3	AC115577	AC115577 Dictyoste
19	100.5	13.0	577	6	BD108716	BD108716 EST and e
20	100	13.0	76351	2	AC124635	AC124635 Mus muscu
21	100	13.0	171586	2	AC138292	AC138292 Mus muscu
22	99.5	12.9	445	6	BD025717	BD025717 Sequence
23	99.5	12.9	898	9	BC022043	BC022043 Homo sapi
24	99.5	12.9	901	9	AF226046	AF226046 Homo sapi
25	97.5	12.6	12190	1	AE001131	AE001131 Borrelia
26	97	12.6	21170	1	AF169008	AF169008 Borrelia
27	96	12.4	348600	1	AB063521	AB063521 Wigglew
28	95.5	12.4	2308	3	BT003555	BT003555 Drosophil
29	95.5	12.4	64195	2	AC015208	AC015208 Drosophil
30	95.5	12.4	184621	3	AC007440	AC007440 Drosophil
31	95.5	12.4	292919	3	AE003823	AE003823 Drosophil
32	95	12.3	2796	9	BC047097	BC047097 Homo sapi
33	95	12.3	6269	9	AB040962	AB040962 Homo sapi
34	95	12.3	138751	10	AC128666	AC128666 Mus muscu
35	95	12.3	182107	2	AC116694	AC116694 Mus muscu
36	94.5	12.2	30885	1	AE001580	AE001580 Borrelia
37	94	12.2	4239	1	BB062901	BB062901 Borrelia bu
38	94	12.2	10479	1	BBAJ3222	BBAJ3222 Borrelia
39	94	12.2	10845	1	AE001129	AE001129 Borrelia
40	93	12.0	207343	9	AC104389	AC104389 Homo sapi
41	93	12.0	299180	1	AP005080	AP005080 Vibrio pa
42	92	11.9	10185	1	AE004286	AE004286 Vibrio ch
43	92	11.9	73472	2	AC084770	AC084770 Homo sapi
44	92	11.9	201414	2	AC109398	AC109398 Rattus no
45	92	11.9	220718	2	AC130762	AC130762 Rattus no

ALIGNMENTS

AX057345
LOCUS AX057345 447 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 11 from Patent WO0075305.
ACCESSION AX057345
VERSION AX057345.1 GI:12310086
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1
AUTHORS Lalanne, J. L. and Rocher, C.
TITLE Novel candida albicans genes and proteins coded by said genes
JOURNAL Patent: WO 0075305-A 11 14-DEC-2000;
HOECHST MARION ROUSSEL (FR)
FEATURES
source
1. .447
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
1. .447
/note="unnamed protein product"
/codon_start=1
/transl_table=12
/protein_id="CAC22438.1"
/db_xref="GI:12310087"
/translation="MSPDIDNVNLLEEOVELGPKGQIGQKDOVLEKGYGYGTG
FORLLIGYIQELMKFWLHSDIQYNNSSIRNLNLELDMAOISITNGDKVEDYEK
NIKARKNLRIASITKEITWKIDSLDLNLEVGSTLQVSNPDDM"

BASE COUNT 192 a 38 c 82 g 135 t
ORIGIN

Alignment Scores:
Pred. No.: 7,61e-65 Length: 447
Score: 772.00 Matches: 148
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-980-054A-12 (1-148) x AX057345 (1-447)

QY 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAACAAATATGAATTAGGA 60

QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
DB 61 TTTAAAGAGGTCAANTACAGGACACAAAGATCAATATTTAGAGGAGAAAGATATGTT 120

QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
DB 121 TATCAAACTGGATTTCAACGATTTTAAATCATCTGTTATATTCAGAGATTAATGAAATTT 180

QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
DB 181 TGGTTATCCATATAGATCAATATAATAACTCTTCTTCACITTCGGAATCATTGGAATAAT 240

QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
DB 241 TTGGAAGATATTATGGCAAAATTTCTATAACGAATGGAGATAAAGAGTTCAAGATTAT 300

QY 101 GluLysAsnIleLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
DB 301 GAAAAAATATTAAAAAGGCAAGAATAAATTTAAGAGTGATAGTAGTATAACTAAAGAA 360

QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThrLeuGlnVal 140
DB 361 ACTTGGAAATTTGATTCATTGATATTTGGTGGAAAGATAGGTGGAACTTTACAAGTT 420

QY 141 SerGluAsnProAspAspMetTrp 148
DB 421 AGTGAACCCCGATGATATGTGG 444

RESULT 2
AX240813
LOCUS AX240813 447 bp DNA linear PAT 26-SEP-2001
DEFINITION Sequence 51 from Patent WO0160975.
ACCESSION AX240813
VERSION AX240813.1 GI:15797749
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE 1
AUTHORS Roemer, T., Jiang, B., Boone, C. and Bussey, H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 51 23-AUG-2001;
Eli Lilly Pharmaceuticals, Inc. (US)
FEATURES
source
1. .447
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"
193 a 38 c 80 g 136 t
ORIGIN

Alignment Scores:
Pred. No.: 4,37e-64 Length: 447
Score: 764.00 Matches: 146
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.65% Mismatches: 0
Query Match: 98.96% Indels: 0
DB: 6 Gaps: 0

US-09-980-054A-12 (1-148) x AX240813 (1-447)

QY 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAACAAATATGAATTAGGA 60

QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
DB 61 TTTAAAGAGGTCAANTACAGGACACAAAGATCAATATTTAGAGGAGAAAGATATGTT 120

QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
DB 121 TATCAAACTGGATTTCAACGATTTTAAATCATCTGTTATATTCAGAGATTAATGAAATTT 180

QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
DB 181 TGGTTATCCATATAGATCAATATAATAACTCTTCTTCACITTCGGAATCATTGGAATAAT 240

QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
DB 241 TTGGAAGATATTATGGCAAAATTTCTATAACGAATGGAGATAAAGAGTTCAAGATTAT 300

QY 101 GluLysAsnIleLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
DB 301 GAAAAAATATTAAAAAGGCAAGAATAAATTTAAGAGTGATAGTAGTATAACTAAAGAA 360

QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThrLeuGlnVal 140
DB 361 ACTTGGAAATTTGATTCATTGATATTTGGTGGAAAGATAGGTGGAACTTTACAAGTT 420

QY 141 SerGluAsnProAspAspMetTrp 148
DB 421 AGTGAACCCCGATGATATGTGG 444

RESULT 3
AX488762
LOCUS AX488762 447 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 6062 from Patent WO02053728.
ACCESSION AX488762
VERSION AX488762.1 GI:22322774

KEYWORDS Candida albicans
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlson, K.L.
Gene disruption methodologies for drug target discovery
Patent: WO 02053728-A 6062 11-JUN-2002;
Elitra Pharmaceuticals, Inc. (US)

TITLE
JOURNAL
AUTHORS
JOURNAL
FEATURES
source
1. .447
Location/Qualifiers
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

BASE COUNT 193 a 38 c 80 g 136 t
ORIGIN

Alignment Scores:
Pred. No.: 4.37e-64 Length: 447
Score: 764.00 Matches: 146
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 98.65% Mismatches: 0
Query Match: 98.96% Indels: 0
DB: 6 Gaps: 0

US-09-980-054A-12 (1-148) x AX488762 (1-447)

QY 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATATAGATAAGTATTAATTTAGAGAAGAACAAATATGAAATTAGGA 60
QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
DB 61 TTTAAAGAGGTCAATACACAGGACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPhe 60
DB 121 TATCAAACTGGATTTCACGATTTTAAATCAATCGTTATTTAGAGGAAAGAAATTT 180
QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
DB 181 TGGTTATCCCATATAGATCATATATTAATCACTTCTTCACATCGAATCATTTGAATAAT 240
QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
DB 241 TTGGAAAATATTTGGCACAATTTCTATACGAATGAGATAAGAGTTGAAGATTAT 300
QY 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
DB 301 GAAAAAATATTTAAAGCGAAGAAATTAATTAAGAGTGATAGCTAGTATTAACATAAGAA 360
QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThrLeuGlnVal 140
DB 361 ACTTGGAAAATTTGATTCATTAGATAATTTGGTGAAGAAGTAGGTGGAATTTTACAAAGTT 420
QY 141 SerGluAsnProAspAspMetTyr 148
DB 421 AGTGAAGAACCCCGATGATATGTGG 444

RESULT 4
LOCUS AX057353 326 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 19 from Patent W00075305.
ACCESSION AX057353
VERSION AX057353.1 GI:12310094
KEYWORDS
SOURCE
ORGANISM
Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
1 Lalanne, J.L. and Rocher, C.

TITLE
JOURNAL
AUTHORS
JOURNAL
FEATURES
source
1. .1104
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="XIV"

Novel candida albicans genes and proteins coded by said genes
Patent: WO 0075305-A 19 14-DEC-2000;
HOECHST MARION ROUSSEL (FR)

FEATURES
source
1. .326
Location/Qualifiers
/organism="Candida albicans"
/mol_type="genomic DNA"
/db_xref="taxon:5476"

BASE COUNT 146 a 27 c 53 g 100 t
ORIGIN

Alignment Scores:
Pred. No.: 4.58e-45 Length: 326
Score: 562.00 Matches: 107
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.07% Mismatches: 0
Query Match: 92.80% Indels: 0
DB: 6 Gaps: 0

US-09-980-054A-12 (1-148) x AX057353 (1-326)

QY 5 AspIleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGlyPheLysGluGly 24
DB 2 GATATAGATATGATTTAAATTTAGAGAAGATCAATATGAAATTAGGATTTAAAGAGGT 61
QY 25 GlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThrGly 44
DB 62 CAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTTATCAAACTGGA 121
QY 45 PheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheThrLeuSerHis 64
DB 122 TTTCAACGATTTTAAATCAATCGTTATTTCAAGAAATTAATGAAATTTGGTTATCCCAT 181
QY 65 IleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
DB 182 ATAGATCATATATTAATCACTTCTTCACATCGAATCATTTGAATAATTTGGAAGATAT 241
QY 85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
DB 242 ATGGCACAATTTCTATACGAATGAGATAAGAAAGTTGAAGATTATGAAAAAATATT 301
QY 105 LysLysAlaArgAsnLysLeuArg 112
DB 302 AAAAGGCAAGAAATTAATTAAGA 325

RESULT 5
LOCUS SCYNL260C 1104 bp DNA linear PLN 11-AUG-1997
DEFINITION S.cerevisiae chromosome XIV reading frame ORF YNL260C.
ACCESSION SCYNL260C
VERSION 271536.1 GI:1302312
KEYWORDS
SOURCE
ORGANISM
Saccharomyces cerevisiae (baker's yeast)
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.

REFERENCE
1 Sen-Gupta, M., Gueldeener, U., Beinbauer, J., Fiedler, T. and Hegemann, J.H.
Unpublished
MIPS.
Direct Submission
Submitted (29-APR-1996) Data collected by MIPS on behalf of the European yeast chromosome XIV sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: News@mips.emblnet.org

FEATURES
source
1. .1104
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/db_xref="taxon:4932"
/chromosome="XIV"

```

CDS      complement (320...916)
          /note="ORF YNL260C"
          /codon_start=1
          /protein_id="CAA96167.1"
          /db_xref="GI:1302313"
          /db_xref="SWISS-PROT:P53846"
          /translation="MVRNRFIRKMKKNLFSKNHLSYLKSKWKVKITQIKQKWDPNLNL
          LEEQYQEGLEGQENIKQSLQKQGLQVGFQFTLLQMEGLCDVIESYGLHSP
          TLEKIHRTILMKGLKNNDESVMERVLIKKPKFTILITLHRLVKDKSTPTV
          TFEVFDVSRALAGIRGVENEDIAKNTKQNOAQS"
BASE COUNT      342 a 206 c 157 g 399 t
ORIGIN
Alignment Scores:      2,72e-12      Length:      1104
Pred. No.:      223.00      Matches:      47
Score:      58.78%      Conservative:      30
Percent Similarity:      58.78%      Mismatches:      40
Best Local Similarity:      35.88%      Indels:      14
Query Match:      28.89%      Gaps:      3
DB:      8

US-09-980-054A-12 (1-148) x SCYNL260C (1-1104)
QY      4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
DB      808 ATGGATTTTGATTAATTACTTAACCTAGAGACAGATTATACCAAGAGGTTTCTGGAG 749
QY      24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThr 43
DB      748 GGCCAGATGAGATATAAAGCAGAGCTTCTAGAGGTAAACAGATATGGTTTACAGTG 689
QY      44 GlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
DB      688 GGATTCCAAGATTACGCTCTCGGACAAATGGAGGTTTGTGATGTT----- 638
QY      64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnLeuGlu 82
DB      637 ---ATAGATCGTATGCGATGTCATAGTCCAACTAGAGAAAGATATACATACCATTCGT 581
QY      83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
DB      580 ACTTTAATGAAGGTTTGAANAATGAATACGATGATGAAGTGTGATGAATTTGAGAGG 521
QY      103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
DB      520 GTGTTAATAAATTGAAATAATAGTTAGAACCATT----- 485
QY      123 LysIleAspSerLeuAspAsnLeuValLysGlu 133
DB      484 ---CTAATAACTCTTCATCGATTAGTTAAGAT 455

RESULT 6
YSCATX1HF      2961 bp      DNA      linear      PLN 06-JUN-1995
LOCUS      Saccharomyces cerevisiae metal homeostasis factor (ATX1) gene,
DEFINITION      complete cds.
ACCESSION      L35270
VERSION      L35270.1 GI:530995
KEYWORDS      metal homeostasis factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
ORGANISM      Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 2961)
Lin,S.J. and Culotta,V.C.
The ATX1 gene of Saccharomyces cerevisiae encodes a small metal
homeostasis factor that protects cells against reactive oxygen
toxicity
Proc.Natl. Acad. Sci. U.S.A. 92 (9), 3784-3788 (1995)
JOURNAL      95249558
MEDLINE      7731983
PUBMED
COMMENT      source text: Saccharomyces cerevisiae DNA.
FEATURES      Original
Location/Qualifiers

source      1..2961
            /organism="Saccharomyces cerevisiae"
            /mol_type="genomic DNA"
            /db_xref="taxon:4932"
            446..667
            /gene="ATX1"
            446..667
            /genes="ATX1"
            /function="protects against oxygen toxicity"
            /codon_start=1
            /product="metal homeostasis factor"
            /protein_id="AAC37428.1"
            /db_xref="GI:530996"
            /translation="WAEIKHYQNVMTCTGCGAGVKNVLTLEPDVSKIDISLEKQL
            VDVTTLPYDFILELIKTKGKVRSGKQL"
BASE COUNT      988 a 466 c 554 g 953 t
ORIGIN
Alignment Scores:      8,22e-12      Length:      2961
Pred. No.:      223.00      Matches:      47
Score:      58.78%      Conservative:      30
Percent Similarity:      58.78%      Mismatches:      40
Best Local Similarity:      35.88%      Indels:      14
Query Match:      28.89%      Gaps:      3
DB:      8

US-09-980-054A-12 (1-148) x YSCATX1HF (1-2961)
QY      4 IleAspIleAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
DB      969 ATGGATTTTGATTAATTACTTAACCTAGAGACAGATTATACCAAGAGGTTTCTGGAG 1028
QY      24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThr 43
DB      1029 GGCCAGATGAGATATAAAGCAGAGCTTCTAGAGGTAAACAGTATGTTTACAGTG 1088
QY      44 GlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
DB      1089 GGATTCCAAGATTACGCTCTCGGACAAATGGAGGTTTGTGATGTT----- 1139
QY      64 HisIleAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnLeuGlu 82
DB      1140 ---ATAGATCGTATGCGATGTCATAGTCCAACTAGAGAAAGATATACATACCATTCGT 1196
QY      83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
DB      1197 ACTTTAATGAAGGTTTGAANAATGAATACGATGATGAAGTGTGATGAATTTGAGAGG 1256
QY      103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
DB      1257 GTGTTAATAAATTGAAATAATAGTTAGAACCATT----- 1292
QY      123 LysIleAspSerLeuAspAsnLeuValLysGlu 133
DB      1293 ---CTAATAACTCTTCATCGATTAGTTAAGAT 1322

RESULT 7
YSCATX1V/c
LOCUS      S.Cerevisiae DNA region from chromosome XIV, left arm.
DEFINITION      X96722
ACCESSION      X96722
VERSION      X96722.1 GI:1255956
KEYWORDS      ATX1 gene; MPA43 gene; NR1 gene; ORC5 gene; RAD50 gene; RPA49
            gene; SIP3 gene; su11 gene.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
ORGANISM      Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
1
Sen-Gupta,M., Guldener,J., Beinbauer,J., Fiedler,T. and
Hegemann,J.H.
Sequence analysis of the 33 kb long region between ORC5 and SU11
from the left arm of chromosome XIV from Saccharomyces cerevisiae

```


JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL	Yeast 13 97377992 9234673 2 (bases 1 to 33016) Hegemann J.H. Direct Submission Submitted (21-MAR-1996) J.H. Hegemann, Institute of Micro & Molecular Biology, Justus-Liebig-Univ. Giessen, Frankfurt Str. 107, D- 35392 Giessen, FRG Location/Qualifiers 1. 33016 /organism="Saccharomyces cerevisiae" /mol_type="genomic DNA" /strain="S288C" /sub_strain="FY1679" /db_xref="taxon:4932" /chromosome="XIV" /map="132kb, left arm" /clone="cosmid 14-4c pou" 223. .1662 /gene="ORC5" 223. .1662 /gene="ORC5" /note="ORF N0834" /codon_start=1 /protein_id="CAA65483.1" /db_xref="GI:1255957" /db_xref="SWISS-PROT:P50874" /translation="MNVTPEVAFREYQTNCLASYISADPDITPSNLILOQYSGTGKT YTKKYNANLNHAWLVELVSKPLQAIARTVOYKLTLPYNIPTIDPDVLP EPELLKTNLHIFVQESLEQCLFLLDGESLQDLDAALFNKYIKLNELLPKDS KINIKFYLMETSLQRYSTHPIPTVFPYRYNDVSTILVMSRCGELMEDSLRKR IIEQITDCTDQFQVANAHLIHOAFHYNCTGNDIPALNDIDFKWPYVSRITKE NIEPLALYKAIFLSTDDNLSENGOGESAITNRDLDLNSDTYDLSIISKVLLIA SYISYLEPRYDASIFSRKTIILQGRAAYGRKKEVNPYLPQSLFAIRLAIPOA IFPQGRASLSALREESLMKANIEVFQNLSELHTKLAIITNNKNDIYLSFKWRW KVNVPWEILKEISSEVHFNISDYFSDIHE" complement (1982. .2578) /note="ORF N0838" /codon_start=1 /protein_id="CAA65484.1" /db_xref="GI:1255958" /db_xref="SWISS-PROT:P53846" /translation="WVRNRIRKWKKNLFFKSNHLSYLSKWKVKITQIQIMDFNLIN LEEYQGFLEGGNNTKQFLGKQYGVQFRTLLGQMEGLCDVIESYGLHSP TLEKNHIRTLMGLKMNNDSEVMFERVLIKLNKFTITILLRLVYDKRTPTV TFEVPDYSRAIAGEIRGFVNEIDAKNKTQKQQAQSM" complement (2767. .2988) /gene="ATX1" complement (2767. .2988) /gene="ATX1" /note="ORF N0840" /codon_start=1 /protein_id="CAA65485.1" /db_xref="GI:1255959" /db_xref="SWISS-PROT:P38636" /translation="MAEIKRHFQNVMTCSGCCSAGVNVKLTLEPDPVSKIDISLEKQL VDVYITLPDYFILEKIKTKTGEVRSKQL" complement (3232. .5496) /note="ORF N0842" /codon_start=1 /protein_id="CAA65486.1" /db_xref="GI:1255960" /db_xref="SWISS-PROT:P53847" /translation="MESLFFPKNGEITIRELLKDPILKNDKSRNGSELELDSLLLOR EAILANELNLDNLTFLNLKEVNTNLNILENCYYSQSLRKKRNNAAYLKQSF NFQGISYVTDTLHLELYTYKILITNGFWKITENSIOFTPTVEWKGDKVHIEYDTFM DFVAQYPPKSGLDNQAMFLDMTSDAQEQVRAKLNTIMKEYNLSRIYSMINKNSIF ISGKEISYENENILVFSKSSSHGOHCYSTVLTSEAVCDPMLDGLAFDRKRLTSYEL GPLENTETKFKVNNASIIISLSPLAKNLVNLNKLTLVAKSEVNTNTHSGKEIQ DLLMKNQYLLNDKVLSIESIRSFEDPKSKQNLVEVLTNTNTNTHSGKEIQ NDSVDQNEKELHNAVSKDDNVEVEDDADWGDENIDNIDDEEKTNOKEKEPPE BENAWDEAWAIDENITDDASLENGEHLKAHDVGLDKDHIETVQLPLFLAISQNFKS	SPADSHVDEQYAYKYNLLQTSYAMACTANFSHNWCQLYVDMRYLIERDEKLYRIKEL TRNLLKTLNMYRIYVQDLIRHQLTREPRENERNPSWDIAIEKLLYILKNIIVRPLQKI RGEESRYLLSFLNFDYQVTKBILKQWILSEVNSNGLSELVLSLVEINVTLOQLAKE PSYKMRKXKATMGKFLPLHLKEINEMFYNGDFYLFATDELQIOWIELLPADTFLRRNA IDDIYIEIRGTALDD" complement (5754. .9443) /gene="SIP3" complement (5754. .9443) /gene="SIP3" /note="ORF N0844" /codon_start=1 /protein_id="CAA65487.1" /db_xref="GI:1255961" /db_xref="SWISS-PROT:P38717" /translation="MSVHGEDDKKQLRLISVAFKEASIDSPSRASVNFOTRVDAL EDWIEKTVDFDQKVKVDFSRRAKFTILSOLLEPPALLSNFVSNQSFTRLIDSF NKDIYDFSMKLOIVKGGDSSHTALLEMTTAIEPYRNVRNKFQYQKIDYSMLASJ QAIRSFTLSPESSIKDALQFVQKNYKASLDLISAIKSLSLDKFILESMKVL KSRISFITKDSGRKIDLSPCINEYLDNVAIWENSIEGSKVLDLISNAKQAYITLL KRITPSSDTSYNIIRSIHSSKLLSDQTPVPKSPKSGMLYMKTOGKPTREIVWRW CFLKNAVFGMFLTSPKTVBETDKGVFLTVRYDPEEDRKFCEVKIFGNKVTEAH DNMXDITLAVFQTSNYLDLKSWLJAPKATKYVMSIQHDSLEYLAEAFKFSKFEFA SSTTSIDQITTFKTESLYETLNCISIEYDITLGEKKVQFQMPPTPISTMTQ LAILNFLTGGWFFNANLWGTDDWSEYITLPGKGGKSSLLITDKRLPIRNT IYPQYSNELKVDLQFKSLVFPDQLEKFEELLFLKFEALGNPKKQKFSATCF TKDYICYMNSMEFICLKISLSEIVSEADRSSKTKLYDASGLQKAIYLFSDYK LIASKLOYLLENKAIKNPNSNEILVQPEOMEKQEQKQEBLYIEQENSFDRKATS VSKILSRVTFWEMSDASTLNLRLKQLQTEYSITYNHEYESKGLAHILFGDKSNA FKCLFLARKDGEHGGKPYNKNDINGKSOALVRKIPRLDNTGNTLANTKYHROKES MFAFORLILVYDKNTYEDUDPPFVKVPFCHLLKLSIKFVITSEYVDNHELEKLN MTASSSHLVLYKLTIDSRQTLIEKLSAELICQTLWAKFAHSEFLIRKRLRYLL EKIGHGVKTIKAIKGLGILVSNKSEBPATKNGSKESQMSQYDIRYSTCILFLV ILKMWRYNTLTFVFRILIGLLCAKFSRINRMVVGILLASINILLSEKASVP YWSIKRAEKLPHDRLGSKDFTMORAIYISDLSLSQSVSPNNPIFEKFSNDNFND YOYSRTRKOLAMRNELLIELLLODMEKQLVHDDYKEFLLSEVNKCSMVSIEMTDLW FNDTQLQVYCSICNEELEKLRPPII" 9626. .12220 /note="ORF N0848" /codon_start=1 /protein_id="CAA65488.1" /db_xref="GI:1255962" /db_xref="SWISS-PROT:P53848" /translation="MINCKVIENRTRRPPPLAKLVKVIASSLINKTSQSLPLAMSKL FETVNSARSHVPLGMDRYVHIKLEMTYLVGDSWNQLMPQKCLLSLDMGTFDSKA ADDDKLSUNYAVISRDLTNFVKKKNGWSVNLNKSQFVMDKISGVECLNLEVQA DTHIRSDIESCIIIOQERGNPESQFVDVIRSELKMLTILIGVTFERLKKQVITLDIK LWPWKAEPPVPVQGIIDNVKVFESNPKTVREALVESVAVIAHEYFKQFPDPLV VKVLKNAITATEGVGVCIRPREIAIMWNIPLYSIHESDIKFLSSQSQPTIEGK NTWKRAFLAFGSGNIGDRFKIQMALQQLLSREKTVKLRNISTIFESEPMYFKDQTPFMN GCVEVETLLTPSEILLKCKHIEYELQVXHFDPNGPRTDLDIVMLFASAGEDIIVNE PLPYRSGVEPRFLKFTATKDEFTGENTRITVSPYIMAFINQKLDKQHDDETLAKLV IESQLNDIILCKDALYHESVIVDGGCTRPNSIQASBEEIIRRSIPILIKAIRST ELPQDKVILSDITYRSNAKEAIVGVDIINDISGLDFSNMFVAIENPETCYTILSH TGDISTNMLAHENFALGDSIQOEFVNTDIQQLDLDKDKTLVIRNVQIEGRYI KALDGVKWEQIILIDPLGFGAKTQKQNLQIIRHPIILKNYSFTMNSNNQSVTVNLRNM PVLGSPRKKFIGHITKDVAQDRDPATGAVVASCIGFGSDVMRVHDVKNCKSIIKLA DAIKKLE" complement (12452. .12913) /note="ORF N0852" /codon_start=1 /protein_id="CAA65489.1" /db_xref="GI:1255963" /db_xref="SWISS-PROT:P53849" /translation="MSQKACYCGKIGHLAEDCDSEELCYNCKNGKPGHVQDCTMPRTV EFKQYCNCGTGEHVRSECTVQRCNCQNGTGHLSRECEPKKTSRSKSKYKCGGPNH MAKDCMKGEDISGLKYCTCGQAGHMRDQCDRLCLNCNCTGHIHSDCPAA" complement (13166. .14371) /note="ORF N0856" /codon_start=1 /protein_id="CAA65490.1" /db_xref="GI:1255964" /db_xref="SWISS-PROT:P53850"	gene CDS gene CDS gene CDS gene CDS
--	--	---	--

```
/translation="MVGFLGGINVRKGVSTTKKSGDNLMLMKRQCKHDIHRESD
DLSDHDAFSPKRSKIDIEDIEIVKSLSTVATFDKLSRFPNSVQAKNAALRG
KEKEKVVSIPLIQNLKNEIDIEIKCNLNNLLDKLLLEALSSEVDNQIFSSFP
EDKLSLOSQSSKEQIKKQVREYMSKFLPMLFSDLLTEVSPFPIVNDILE
GKISVYVFEAKNAFKSKAYLVSDPERKLNANKFTAGFYGLKQRLRVGEIAKRYK
RALITHQFATLKWGITDFCYVLAPELTISFCIYQNLNLSKSSCTKPKHQQLN
EKREYIDPELRMLAYDLLEDIVYGIIVADSDPIEQWEAAIEEDRLRELKLDVHNYSS
RWRLDTHD"
CDS
15141..16409
/notes="ORF N0860"
/codon_start=1
/protein_id="CAA65491.1"
/db_xref="GI:1255965"
/db_xref="SWISS-PROT:P53851"
/translation="MSTGAVDILNOKTITSEVAASVTSKYLSQSTFKGNTSHIEDKR
FHVSRSHSRFTPTIPNEILSLKHFVSGSSMAYSDGSLTVFWFKDASFORSVE
VFIPOCCSDKLATLSNPTSLNQIAVVSNSSEISLILLINEKSLTASKLRTLSLGSK
TKVNTCLDPLGNWLLATKSKYLFVQKDHSSVCSLINSIDSDNDVYSLAWS
NGSHIFIGFGSYLAALKAKHGLEVCTKIKATGPITEIKMDPWRGNFTIGSIDGN
CYVWMSKSLCBLIINDLSAVTTLDVCHLGLICTDEDMVYFYDLNSGLNLSKSKS
LANYIKTDPVKFTPKSNTGKRDTPSLNHFVKNKLIITWKMDPMTWIEKRRKN
NGGNNHNRKTSKNTDRIGKDRSPNSKK"
complement(16564..17409)
/notes="ORF N0864"
/codon_start=1
/protein_id="CAA65492.1"
/db_xref="GI:1255966"
/db_xref="SWISS-PROT:P36528"
/translation="MKVNLMLKRGATATATATASSAPPKIKGVLLSRIPILKSELNEL
EKRYEYQSELEKRLMTFPAYFYKGTGVAEHKFLSKQKIPSKNGIWFPRGIPDI
KHGRESTQKQLSDSTVAFNNKQKQSDVNRVPIPNDRTEADRSRDMKSLER
QLSRTLVLVQKQSKTWKPNPLDSEKPLHVAENELKLSGQIYTWVSATPIG
VLQDERNRTAEFIVKSHILLAGKFDLVASKNDAPFAMLTKEISEYVVPKDYFNKTEF
LIADN"
complement(17711..19438)
/gene="NRD1"
complement(17711..19438)
/gene="NRD1"
/notes="ORF N0868"
Alignment Scores:
Pred. No.: 1.23e-10 Length: 33016
Score: 223.00 Matches: 47
Percent Similarity: 58.78% Conservative: 30
Best Local Similarity: 35.88% Mismatches: 40
Query Match: 28.89% Indels: 14
DB: Gaps: 3
US-09-980-054A-12 (1-148) x SCCHXIVL (1-33016)
Qy 4 IleAspLeuAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGlu 23
Db 2470 ATGGATTTCGATAATTACTAAACCTAGAGAACACAGTATTACCAAGAGTTTCTCGAG 2411
Qy 24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGlyTyrGlnThr 43
Db 2410 GGCAGAAATGAGATAAAGACGAGAGCTTCTAGAGGTAAACAGATGTTTACAGTG 2351
Qy 44 GlyPheGlnArgPheLeuIleGlyTyrIleGlnLeuLeuMetLysPheTrpLeuSer 63
Db 2350 GGATTCACAAAGATTACGCTCCTGGGACAAATGGAAGTTTGTGTGATGTT----- 2300
Qy 64 HisLeuAspGlnTyrAsn---AsnSerSerSerLeuArgAsnHisLeuAsnLeuGlu 82
Db 2299 ---ATAGAATCGTATGATGTCATAGTCCAACTAGAAAGAAATATACATACATTCGT 2243
Qy 83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
Db 2242 ACTTAAATGAAGGTTTGAAATCAATCAACGATGATGAAGTGTGATGGAATTTGAGAGG 2183
Qy 103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrp 122
Db 2182 GTGTAAATAAATTGAAAAATTAAGTTTAGACACCT----- 2147
```

```
Qy 123 LysIleAspSerLeuAsnLeuValLysGlu 133
Db 2146 ---CTAATAACTCTTCATCGATTAGTTAAGAT 2117
RESULT 8
SPCC191
LOCUS SPCC191 25000 bp DNA linear PLN 29-MAR-2000
DEFINITION S.pombe chromosome III cosmid c191.
ACCESSION AL049644
VERSION AL049644.1 GI:4678667
KEYWORDS acetyl-coenzyme a synthetase; AMP-binding, AMP-binding; cycl1;
cytochrome c; glutathione s-transferase; Glycosyl hydrolase; invl1;
invertase; LTR; pseudogene.
SOURCE Schizosaccharomyces pombe (fission yeast)
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 25000)
AUTHORS Lyne M., Rajandream M.A., Barrell B.G. and Volckaert G.
TITLE Direct Submission
JOURNAL Submitted (24-MAR-1998) European Schizosaccharomycetes genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk
and atholieke Universiteit Leuven, Faculty of Agricultural and
Applied Biological Sciences, Laboratory of Gene Technology,
Kardinaal Mercierlaan 92 Blok F, B-3001 Leuven, Belgium
COMMENT
Notes:
Details of yeast sequencing at the Sanger Centre are available on
the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_pombe/)
Protein coding regions (CDS) have been predicted with the help of
computer analysis using the Genefinder program in PomBase (an ACEDB
database) with additional predictions for the branch-acceptor sites
supplied by the program Sp3splice. CAUTION: It is possible that for
any individual CDS we may have underestimated or overestimated the
number of introns/exons or we may not have chosen the correct
splice donor/acceptor sites.
CDS are numbered using the following system eg SPAC5H10.01c.SP (S.
pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
(complementary strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid c191 overlaps c417 at the left end and c1450 at the right
end.
FEATURES
source
Location/Qualifiers
1..25000
/organism="Schizosaccharomyces pombe"
/mol_type="genomic DNA"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/map="I"
/clones="cosmid c191"
1..4556
/misc_feature
/notes="nominal overlap with SPCC417 AL035076 S.pombe
chromosome III cosmid c417."
2268..2804
/gene="SPCC191.01"
/notes="synonym: SPCC417.13"
2268..2804
/gene="SPCC191.01"
/notes="SPCC191.01, len:178"
/codon_start=1
/label="SPCC191.01"
/product="hypothetical protein"
/protein_id="CAB41047.1"
```


(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologues of the coding regions were searched against NCBI NonRedundant Protein database with BLASTP. ESTs represent the identified cDNA sequences using BLASTN with the corresponding DBJ accession no. and RGP clone ID.

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, putative- and -like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence.

The orientation of the sequence is from T7 to SP6 of the PAC clone. This sequence of P0681F05 clone has an overlap with OSUNBA0007H12 (DBJ: AF004990) clone at 5' end and with OSUNBB0044C15 (DBJ: AP005950) at 3' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://rgp.dna.affrc.go.jp/GenomeSeq.html>.

Location/Qualifiers

FEATURES

source

1. 144741

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/chromosome="7"

/clone="P0681F05"

3263. 3469

/gene="P0681F05.1"

3263. 3469

/gene="P0681F05.1"

/note="hypothetical protein"

similar to Oryza sativa chromosome7, OJ1656_E11.24"

/codon_start=1

/protein_id="BAC45137.1"

/db_xref="GI:27261021"

/translation="MADGYRRVLTAASTLPLVACSGAGMCQVPAARGGRLFGVA

TGGDRLRCGWPRKAVESAFOMAD"

complement(join(526

7335. 7831,7975..7985))

/gene="P0681F05.2"

complement(join(6526

7335. 7831,7975..7985))

/gene="P0681F05.2"

/note="hypothetical protein"

predicted by EGENSEH etc."

/codon_start=1

/protein_id="BAC45138.1"

/db_xref="GI:27261022"

/translation="MTWKPSLHQSQMDPGSAGGARNDWCMCEFTSRRYIAEK

SRAAPGSSSHGLASCEMPPCKMQGPVLIEMFKPRGTGTYIGSNPLSGWEQVQ

NRTPTRFWS"

complement(join(13012

complement(join(13012

/gene="P0681F05.3"

/note="hypothetical protein"

predicted by GeneMark.hmm etc."

/codon_start=1

/protein_id="BAC45139.1"

/db_xref="GI:27261023"

/translation="MTDAGDDEPESEERAAARDSDDESSAARLASPPPELGL

GGLSQPRLNHLGLVRLNGLGEGNQDPT"

14320. 16593

/gene="P0681F05.4"

14320. 16593

/gene="P0681F05.4"

/note="probably inactive due to stop codon(s) in CDS

pseudogene, transposase"

/pseudo

Join(19824. 19909,20071. 20178,20294. 20405)

gene

gene

CDS

/gene="P0681F05.5"

Join(19824. 19909,20071. 20178,20294. 20405)

/gene="P0681F05.5"

/note="hypothetical protein

predicted by GeneMark.hmm etc."

/codon_start=1

/protein_id="BAC45140.1"

/db_xref="GI:27261024"

/translation="WGHEELVAGSGDAEAGDQRPRVAARRAATPGGDEARTSN

GQSTIRRRPVAENADERRSGAENAVGADERRRSVAEVADEMTAGRGGGRG"

Join(23817. 24701,24780. 24953)

/gene="P0681F05.6"

Join(23817. 24701,24780. 24953)

/gene="P0681F05.6"

/note="contains ESTs

D41499(S4032).AU034265(E3758).AU033141(S4032)"

/codon_start=1

/product="putative thiamine biosynthetic enzyme"

/protein_id="BAC45141.1"

/db_xref="GI:27261025"

/translation="MATTASILLKTSFAGARLPAAARNPTVSVAPRTGCAICNLSISS

SSTPPYDINAIRFSPIKESIVSREMTYMDITVADTVVVGAGAGLSCAYELS

KDPSVNAVIRQSVPGGANLGGLESAMVVRKPAHLFLDELGVAYDEQDYVVIKH

ALPFTSMRLRLAPNKLLENVAVEDLIVKEGTVGVVTVNVALVSNHNTQSCMDP

NVMSRVVSSCGHDGPGFGVKKLQDGMIDAVPGMRALDMNTAEIEIVLTREVY

FGMIVTGEVABIDGAPRMGPTFGAMMTSGQAAHLALKALGRPNADITGIKAAAAA

AHPELILASKDDGEIVDA"

complement(join(27163. 27332,27442. 27529))

/gene="P0681F05.7"

complement(join(27163. 27332,27442. 27529))

/note="This category is not included in IRGSP standard.

hypothetical ORF

predicted by GENSCAN"

Join(28697. 28907,28986. 29086,29178. 29380,29774. 30188)

/gene="P0681F05.8"

Join(28697. 28907,28986. 29086,29178. 29380,29774. 30188)

/gene="P0681F05.8"

/note="contains ESTs C72892(E2426).AU172540(E2426)"

/codon_start=1

/product="putative xyloglucan endotransglycosylase"

/protein_id="BAC45142.1"

/db_xref="GI:27261026"

/translation="MGSIGRRPWWGLTAAMIFAVAVCGFCFSGASAAAAAFTFGDNF

ETGAEDHVKSADQTYLYLDNKTGYGQFKERYLFGWFSMNLKLAGNSAGVYTA

YVCSLDNPKAIPORDELDFLGNRTGEEPIQTIVIRSVGSGRENHSLWEDPTADFH

SYSLWNPQAIYFFVDKPIREYNSDKPTEFFPIMKPMYVSSIWNADDWATRGLE

KTDWTKAPFISSYRDTADACSWGTAAASPSCAASTGNSMWDQPPAWALDAGQREDS

AWARNLVIYDCCDRKPPSPPECLLRTTSS"

complement(join(32633. 32677,32784. 32837,32929. 33140,

33226. 33262))

/gene="P0681F05.9"

complement(join(32633. 32677,32784. 32837,32929. 33140,

33226. 33262))

/gene="P0681F05.9"

/note="contains ESTs

AU093360(C1467).AU075849(S20385).D28232(C1467),

AU075848(S20385)"

/codon_start=1

/product="translational initiation factor eIF1"

/protein_id="BAC45143.1"

/db_xref="GI:27261027"

/translation="MSDLDIQIPTAFDPFAEANAGDSGAAGSKDYVHVRIQQRNGRK

SLTWTGGLAKKFSYNNKILDKKFCNCNGTVQDPELGQVQLQDQRKNVSNFLVQA

GIVKKEHIIKHGF"

Join(36723. 36827,37444. 37566)

/gene="P0681F05.10"

Join(36723. 36827,37444. 37566)

/gene="P0681F05.10"

/note="This category is not included in IRGSP standard.

hypothetical ORF

predicted by GlimmerM"

38236. 38406

gene

gene

gene

gene

79 AsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGly- 93
Db 24404 CAAAAGTCGAGGATAGCTAGAAAACCTTAACCTGGGTAAACAACAGCGGAGAAAACA 24463
Qy 93 ----- 93
Db 24464 GAGGAGTCGCGCTACTGCCCCAACACGCGCTGTAGAAAACATGCTCTACCTCTTT 24523
Qy 94 ----- AspLysGluValGluAspTyr 100
Db 24524 GCGCTGTCGAGACCGACACGCTAGCGAAAGAAAACACGAGCGAGCGTGAATGATTT 24583
Qy 101 GluLysAsnIleLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
Db 24584 GATGATCGCATGAGGCGCGGAGGAGAAAGCAAGATGTG- 24625
Qy 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThrLeuGlnVal 140
Db 24626 -----GAGAGGATGCTGGGGAGGTCTCAGAAAAGAGGAGCG 24664
Qy 141 SerGluAsnPro 144
Db 24665 AATACGAATCCT 24676
RESULT 12
LOCUS AF360255
DEFINITION Arabidopsis thaliana unknown protein (At2g20830) mRNA, complete cds.
ACCESSION AF360255.1 GI:13430685
VERSION AF360255.1
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1 (bases 1 to 1904)
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Arabidopsis Full Length cDNA Clones
TITLE Arabidopsis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1904)
Yamada,K., Liu,S.X., Sakano,H., Pham,P.K., Banh,J., Chung,M.K.,
Goldsmith,A.D., Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L.,
Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J.,
Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Kim,C., Lam,B.,
Lin,J., Miranda,M., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T.,
Satou,M., Seki,M., Shinn,P., Southwick,A., Shinozaki,K.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Direct Submission
TITLE Submitted (13-MAR-2001) Plant Gene Expression Center, 800 Buchanan
JOURNAL Street, Albany, CA 94710, USA
COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.
The Salk, Stanford, PGECC (SSP) Consortium members carried out the
sequencing and annotation of the RAPL cDNAs: Yamada,K., Liu,S.X.,
Sakano,H., Pham,P.K., Banh,J., Chung,M.K., Goldsmith,A.D.,
Lee,J.M., Quach,H.L., Toriumi,M., Yu,G., Bowser,L., Chen,H.,
Cheuk,R., Jones,T., Karlin-Neumann,G., Kim,C., Lam,B., Lin,J.,
Miranda,M., Nguyen,M., Palm,C.J., Shinn,P., Southwick,A.,
Davis,R.W., Ecker,J.R. and Theologis,A.
Yamada,K. (SSP/PGECC) and Seki,M. (RIKEN GSC) contributed equally to

this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGECC)
contributed equally to this work as PIs.

Annotation is based on the January 2002 version of the Arabidopsis
genome submitted to Genbank.

FEATURES

Location/Qualifiers
source
1..1904
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="RAF109-16-J24 (R09231)"
/notes="This clone is in a modified pBluescript vector
(pUC-1) as a BamHI/XhoI insert."
ecotype: Columbia
1..1904
/genes="At2g20830"
1..832
/genes="At2g20830"
833..1726
/genes="At2g20830"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AAK25965.1"
/db_xref="GI:13430686"
/translation="MLREMLGCCVYISEARNKTALEAIERALKPFPFAIIVNKFDA
AYGRGVTVVSSSLANGSSSLNKNAPVMTALDTINLEHCGSHRPLGVVDHICFPH
LSQTSIEOVSSVANSIAMDIGSTLRVPTVLYGAAKEOCTLDIRKLGKVFANRIGH
EWAGGEDLWVPLKPDAGPOEVSKAGVAVGACGVSNVYPMNSDLKAVRRIARK
TSRGGGLASVQTMALVHGEGVIEVACNLNPSQGVGDEVOGLIERLGREGLLVKGK
YYTDTTPDQIVERYIMDLNNS"
1727..1904
/genes="At2g20830"
misc_difference 1742..1743
/genes="At2g20830"
/notes="compared to genomic sequenc"
/replace="atg"
1889
misc_feature
/genes="At2g20830"
/notes="artifact within poly A tail"
BASE COUNT 523 a 370 c 456 g 555 t
ORIGIN
Alignment Scores:
Pred. No.: 0.214 Length: 1904
Score: 111.00 Matches: 31
Percent Similarity: 46.49% Conservative: 22
Best Local Similarity: 27.19% Mismatches: 53
Query Match: 14.38% Indels: 8
DB: 8 Gaps: 2
US-09-980-054A-12 (1-148) x AF360255 (1-1904)
Qy 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
Db 179 CTCGATTGTTATCGTGGCTTTAGAGGAGACACATGTTTCACAGCGTTTCGATGAGGGTTAC 238
Qy 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGlyTyrGlnThrGlyPhe 45
Db 239 GAAGAAGGCTCTGTGTCGCGGTGCGTGAAGAGCGCTCGTCATCTGGGTTTGAACCTCGGTTTC 298
Qy 46 GlnArgPheLeuIleLeGlyTyrIleGlnGlnLeuMetLysPheTrpLeuSer----- 63
Db 299 GAGACAGCGAGCTGATGGATTCTACAGAGGTTGCTCTGCTCTTTTGGAAATTCAGCTCTC 358
Qy 64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
Db 359 CGTATTGATCTCTACACGCTTCTCTCCTCAGCTCCATAGAGCATCTCAATGATTTCCATGTC 418
Qy 84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
Db 419 TTGCTCGATAAATCCCGCTTTTGGATCCCGAGGACGAGCAAGTAAGAC----- 466

QY 104 IllelylsylsalaarganlystleuArgvalilleAlaserIle 117
 Db 467 -----GGGATCAAGGATGATCTCAGATGCAATTCAGCATC 502

RESULT 13
 LOCUS AC006234 129667 bp DNA linear PLN 11-MAR-2002
 DEFINITION Arabidopsis thaliana chromosome 2 clone F5H14 map mil148, complete sequence.
 ACCESSION AC006234
 VERSION AC006234.4 GI:20197676
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 129667)
 AUTHORS Lin, X., Kaul, S., Shea, T.P., Fujii, C.Y., Shen, M., VanAken, S.E., Barnstead, M.E., Mason, T.M., Bowman, C.L., Ronning, C.M., Benito, M.-I., Carrara, A.J., Creasy, T.H., Buehl, C.R., Town, C.D., Nierman, W.C., Fraser, C.M. and Venter, J.C.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 129667)
 AUTHORS Lin, X.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
 REFERENCE 3 (bases 1 to 129667)
 AUTHORS Town, C.D. and Kaul, S.
 TITLE Direct Submission
 JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@tigr.org
 COMMENT On Apr 18, 2002 this sequence version replaced gi:6598541.
 FEATURES
 source
 1..129667
 /organism="Arabidopsis thaliana"
 /moltype="genomic DNA"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="2"
 /map="mil148"
 /clones="F5H14"
 misc_feature
 1..29862
 /note="overlap with BAC clone F26H11 (AC006264:63359..93220)."
 398..426
 /rpt_family="AT rich"
 repeat_region
 complement(1061..1105)
 /rpt_family="AT rich"
 repeat_region
 complement(1338..1474)
 /rpt_family="AT rich"
 repeat_region
 complement(1487..1891)
 /rpt_family="AT rich"
 repeat_region
 complement(1640..1689)
 /rpt_family="AT rich"
 repeat_region
 complement(1760..1797)
 /rpt_family="AT rich"
 repeat_region
 complement(2893..3840)
 /rpt_family="AT rich"
 gene
 /gene="At2g21045"
 /note="synonym: F5H14.48; contains similarity to senescence associated gene Ntadin from Gi:7594903 (Nicotiana tabacum); contains C2 domain profile (PS50004:PF00168)";
 complement(join(<2893..3015,3089..3142,3253..3323,3412..3485,3566..3643,3731..3840))
 /rpt_family="AT rich"
 mRNA
 /db_xref="GI:20197683"
 /translation="MGMINPVYVIELSDKISKKTIVKHNLPENWEEFKFSVRDP KTVLEFNVYGEKIGKDKCMYVLAKEAPDERKAFTELRKTLGGEGGPGKY RKLVELLYKFTTEEMQAVOKAREGTFVAGGMLVIVHSAEDVEGHNTNYPVHIY FKGEERKTKNVKKNPKWNEEFSEFMLEPEPVHVKHVEVFTSSRIGLLHPKVTFLP FSG"
 CDS
 complement(join(2893..3015,3089..3142,3253..3323,3412..3485,3566..3643,3731..3840))
 /gene="At2g21045"
 /codon_start=1

/product="senescence-associated protein"
 /protein_id="AA015209.1"
 /db_xref="GI:20197696"
 /translation="MYTYSLLNLNLSHCRROTRKRKRKTDHTTEGLMEETKPKTVEDEVV DYTAKGLSTGHRYLDVRTNEEFKSHVEALNI PYMEKTDGRVNPDPFSOVASV CKDDEHLIVACNAGGRSRACVDLLNREGYDHVANMGGSYSAWVDAGFAGDKPPDLKI ACKRPRKEN"
 complement(4388..5541)
 /gene="At2g21040"
 /note="synonym: F5H14.40"
 complement(join(<4388..4658,4742..4831,4920..5018,5109..5156,5264..5541))
 /gene="At2g21040"
 complement(join(4388..4658,4742..4831,4920..5018,5109..5156,5264..5541))
 /gene="At2g21040"
 /codon_start=1
 /product="predicted protein"
 /protein_id="AA015215.1"
 /db_xref="GI:20197703"
 /translation="MQAVQKRAPECTPAAGGMFVIVHSAEDVEGKHTNPVYHLYEKG EERKTKVKNKPKWNEEFSEFMLEPEPIHKMKVYFSSRIVFVLRCA5EAAVY PFEGIKPRGAKLFDGSACSGKGDITTFGDMALLGGMALMLAHDGSLGFLTCHVMYRE SFEEINMAWRAKLVROGAGSSGFWRRFVRELVEQAEVROGAGLSSGSRFVKEPVR QGAGSGVGVNPLK5QGRPSDLRFLKSTSSFFKFDHYKLIK"
 6120..6794
 /gene="At2g21030"
 /note="synonym: F5H14.46; predicted by genscan"
 join(<6120..6214,6312..6441,6549..6794)
 /gene="At2g21030"
 join(6120..6214,6312..6441,6549..6794)
 /gene="At2g21030"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA015206.1"
 /db_xref="GI:20197693"
 /translation="MVKLTCTIILNESHITFASLTGCGNDLKIRFSREMYKNQAOQR WGENYKIVELYNVQNFNRQALQTPARSDDDSTYSKMSARSKDWTPHNFRPPG SVPHFYGGSSNYGPGSYHGPGPMDAARTTTSSRDDPPPSMSNASEMQAEWTEED"
 7487..8921
 /gene="At2g21020"
 /note="synonym: F5H14.6"
 join(<7487..7651,8023..8247,8523..8921)
 /gene="At2g21020"
 join(7487..7651,8023..8247,8523..8921)
 /gene="At2g21020"
 /codon_start=1
 /product="putative major intrinsic (channel) protein"
 /protein_id="AA015199.1"
 /db_xref="GI:20197680"
 /translation="MFGERVIVF5MAISDITQTQTVVSDIENDRSDISQSSVLS GPLSVSVFQKLIGEVGTFSMIFAGCSAIVNETYKPVTLPGIALVWGLVTVTVIY SIGHVSGAHFNPAVSIAFASKPKFPNQHPHYQVHGLMVCYDRQNHVRLVYQFA IKYLRKIAKLPKQRFVFEWILRLIKTILYVHPLIPSMANVHLVSWRELRYRCTTHH TSSISMVYFVSFHITVIFILYVAISDKLIYSINTKCFMLNFI"
 9534..9619
 /rpt_family="AT rich"
 repeat_region
 complement(10013..10754)
 /gene="At2g21010"
 /note="synonym: F5H14.47"
 complement(join(<10013..10504,10581..10754))
 /gene="At2g21010"
 complement(join(10013..10504,10581..10754))
 /gene="At2g21010"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA015200.1"
 /db_xref="GI:20197683"
 /translation="MGMINPVYVIELSDKISKKTIVKHNLPENWEEFKFSVRDP KTVLEFNVYGEKIGKDKCMYVLAKEAPDERKAFTELRKTLGGEGGPGKY RKLVELLYKFTTEEMQAVOKAREGTFVAGGMLVIVHSAEDVEGHNTNYPVHIY FKGEERKTKNVKKNPKWNEEFSEFMLEPEPVHVKHVEVFTSSRIGLLHPKVTFLP FSG"

* NOTE: This record contains 71 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 713: contig of 713 bp in length
* 813: gap of 100 bp
* 844 1529: contig of 716 bp in length
* 1530 1629: gap of 100 bp
* 1630 2354: contig of 725 bp in length
* 2355 2454: gap of 100 bp
* 2455 3192: contig of 738 bp in length
* 3193 3292: gap of 100 bp
* 3293 3991: contig of 699 bp in length
* 3992 4091: gap of 100 bp
* 4092 4823: contig of 732 bp in length
* 4824 4923: gap of 100 bp
* 4924 5649: contig of 726 bp in length
* 5650 5749: gap of 100 bp
* 5750 6467: contig of 718 bp in length
* 6468 6567: gap of 100 bp
* 6568 7279: contig of 712 bp in length
* 7280 7379: gap of 100 bp
* 7380 8056: contig of 677 bp in length
* 8057 8156: gap of 100 bp
* 8157 8881: contig of 725 bp in length
* 8882 9701: contig of 720 bp in length
* 9702 9801: gap of 100 bp
* 9802 10517: contig of 716 bp in length
* 10518 10617: gap of 100 bp
* 10618 11348: contig of 731 bp in length
* 11349 11448: gap of 100 bp
* 11449 12185: contig of 737 bp in length
* 12186 12285: gap of 100 bp
* 12286 13001: contig of 716 bp in length
* 13002 13101: gap of 100 bp
* 13102 13812: contig of 711 bp in length
* 13813 13912: gap of 100 bp
* 13913 14639: contig of 727 bp in length
* 14640 14739: gap of 100 bp
* 14740 15450: contig of 711 bp in length
* 15451 15550: gap of 100 bp
* 15551 16279: contig of 729 bp in length
* 16280 16379: gap of 100 bp
* 16380 17114: contig of 735 bp in length
* 17115 17214: gap of 100 bp
* 17215 17949: contig of 735 bp in length
* 17950 18049: gap of 100 bp
* 18050 18769: contig of 720 bp in length
* 18770 18869: gap of 100 bp
* 18870 19604: contig of 735 bp in length
* 19605 19704: gap of 100 bp
* 19705 20394: contig of 690 bp in length
* 20395 20494: gap of 100 bp
* 20495 21232: contig of 736 bp in length
* 21233 21332: gap of 100 bp
* 21333 22059: contig of 727 bp in length
* 22060 22159: gap of 100 bp
* 22160 22877: contig of 718 bp in length
* 22878 22977: gap of 100 bp
* 22978 23707: contig of 730 bp in length
* 23708 23807: gap of 100 bp
* 23808 24531: contig of 724 bp in length
* 24532 24631: gap of 100 bp
* 24632 25363: contig of 732 bp in length
* 25364 25463: gap of 100 bp

* 25464 26197: contig of 734 bp in length
* 26198 26297: gap of 100 bp
* 26298 27036: contig of 739 bp in length
* 27037 27136: gap of 100 bp
* 27137 27880: contig of 744 bp in length
* 27881 27980: gap of 100 bp
* 27981 28707: contig of 727 bp in length
* 28708 28807: gap of 100 bp
* 28808 29527: contig of 720 bp in length
* 29528 29627: gap of 100 bp
* 29628 30351: contig of 724 bp in length
* 30352 30451: gap of 100 bp
* 30452 31135: contig of 684 bp in length
* 31136 31235: gap of 100 bp
* 31236 31953: contig of 718 bp in length
* 31954 32053: gap of 100 bp
* 32054 32774: contig of 721 bp in length
* 32775 32874: gap of 100 bp
* 32875 33606: contig of 732 bp in length
* 33607 33706: gap of 100 bp
* 33707 34436: contig of 730 bp in length
* 34437 34536: gap of 100 bp
* 34537 35273: contig of 737 bp in length
* 35274 35373: gap of 100 bp
* 35374 36104: contig of 731 bp in length
* 36105 36204: gap of 100 bp
* 36205 36926: contig of 722 bp in length
* 36927 37026: gap of 100 bp
* 37027 37741: contig of 715 bp in length
* 37742 37841: gap of 100 bp
* 37842 38549: contig of 708 bp in length
* 38550 38649: gap of 100 bp
* 38650 39370: contig of 721 bp in length
* 39371 39470: gap of 100 bp
* 39471 40208: contig of 738 bp in length
* 40209 40308: gap of 100 bp
* 40309 41033: contig of 725 bp in length
* 41034 41133: gap of 100 bp
* 41134 41856: contig of 723 bp in length
* 41857 41956: gap of 100 bp
* 41957 42701: contig of 745 bp in length
* 42702 42801: gap of 100 bp
* 42801 43537: contig of 736 bp in length
* 43538 43637: gap of 100 bp
* 43638 44362: contig of 725 bp in length
* 44363 45182: contig of 720 bp in length
* 45183 45282: gap of 100 bp
* 45283 46002: contig of 720 bp in length
* 46003 46102: gap of 100 bp
* 46103 46814: contig of 712 bp in length
* 46815 47643: contig of 729 bp in length
* 47644 47743: gap of 100 bp
* 47744 48473: contig of 730 bp in length
* 48474 48573: gap of 100 bp
* 48574 49392: contig of 719 bp in length
* 49393 49392: gap of 100 bp
* 49393 50130: contig of 738 bp in length
* 50131 50230: gap of 100 bp
* 50231 50961: contig of 731 bp in length
* 50962 51061: gap of 100 bp
* 51062 51791: contig of 730 bp in length
* 51792 51891: gap of 100 bp
* 51892 52809: contig of 718 bp in length
* 52809 52709: gap of 100 bp
* 52710 53414: contig of 705 bp in length
* 53415 53514: gap of 100 bp
* 53515 54347: contig of 733 bp in length
* 54348 54347: gap of 100 bp
* 54348 55080: contig of 733 bp in length
* 55081 55180: gap of 100 bp
* 55181 55889: contig of 709 bp in length

* 55890 55899: gap of 100 bp
* 55990 56695: contig of 706 bp in length
* 56696 56795: gap of 100 bp
* 56796 57117: contig of 722 bp in length
* 57118 57617: gap of 100 bp
* 57618 58350: contig of 733 bp in length.

Alignment Scores:
Pred. No.: 19.2 Length: 58350
Score: 108.00 Matches: 39
Percent Similarity: 51.80% Conservative: 33
Best Local Similarity: 28.06% Mismatches: 41
Query Match: 13.93% Indels: 26
DB: 2 Gaps: 7

US-09-980-054A-12 (1-148) x AC137693 (1-58350)

QY 22 LysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyr 41
Db 9684 GAGGAGGTGAGTCCATTTTAAAAACCAAAANATATGATACAAAGATGAGGGAG 9625
QY 42 GlnThrGlyPheGlnArgPheLeuIlelleGlyTyrIleGlnGluLeuMetLysPheTyr 61
Db 9624 AAATCTTCAGTTAAAGAGATAGATATAATAAAAAACATTCAAAACCTTCAGGAACAATGG 9565
QY 62 LeuSerHisIleAspGlnTyrAsnAsnSerSer----- 72
Db 9564 ATGATGGATGC---ACTATAGAAATGCAAAATGCTCTAGAAAGCTCAGCAATAGAT 9508
QY 73 ---SerLeuArgAsnHisLeuAsnLeuGluAspIleMetAlaGlnIleSerIleThr 91
Db 9507 CCAAGCAGAGAAT-----TCAGAGCTTGAAGACAAAGTTTTCGAATTAACCAATCC 9454
QY 92 AsnGlyAspLysGlu-----ValGluAspTyrGluLysAsnIleLysLys----- 106
Db 9453 ACAAGACAGAGAAAAGAAATAGAAATATGACAAAGCTTCAGAAATCTGGAT 9394
QY 107 -----AlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIle 124
Db 9393 TAGTTTAAACATCAACACCTAGATAATC---AGCATTCTGAGAAAGAGAAATCT 9337
QY 125 AspSerLeuAsnLeuValLysGluValGlyThrLeuGlnValSerGluAsn 143
Db 9336 AAAAGTTTGAAAACATATCCAGGGAATA-----ATCAGGAAGAAC 9295

RESULT 15
AE001581 30651 bp DNA circular BCT 17-MAR-2003
LOCUS Borrelia burgdorferi B31 plasmid cp32-9, complete plasmid sequence.
DEFINITION Borrelia burgdorferi B31
ACCESSION AE001581
VERSION AE001581.1 GI:6382346
KEYWORDS
SOURCE
ORGANISM
Borrelia burgdorferi B31
Borrelia burgdorferi B31
Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia;
Borrelia burgdorferi group.

1 (bases 1 to 30651)
Casjens,S., Palmer,N., van Vugt,R., Huang,W.M., Sutton,G.G., Lathigra,R.,
Rosa,P., Lathigra,R., Sutton,G., Peterson,J., Dodson,J., Haft,D.,
Hickey,E., Gwinn,M., White,O., and Fraser,C.M.
A bacterial genome in flux: the twelve linear and nine circular
extrachromosomal DNAs in an infectious isolate of the Lyme disease
spirochete Borrelia burgdorferi
Mol. Microbiol. 35 (3), 490-516 (2000)

2 (bases 1 to 30651)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
White,O., Dodson,J., Hickey,E.K., Gwinn,M., Peterson,J.,
van Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
Direct Submission
Submitted (11-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
FEATURES
source

3 (bases 1 to 30651)
Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Lathigra,R.,
White,O., Dodson,J., Hickey,E.K., Gwinn,M., Peterson,J.,
van Vugt,R., Palmer,N., Haft,D., Rosa,P. and Stevenson,B.
Direct Submission
Submitted (21-JUN-2002) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Amino acid sequence update by submitter
Location/Qualifiers
1..30651
/organism="Borrelia burgdorferi B31"
/mol_type="genomic DNA"
/db_xref="taxon:224326"
/plasmid="cp32-9"
66..1292
/gene="BBN01"
66..1292
/gene="BBN01"
/note="hypothetical protein; identified by Glimmer2;
putative"
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF07681.1"
/db_xref="GI:6382369"
/translation="MCDLRKTLIDKISSLELYKYSIFRNVIENVAEDCLNGLVLE
SADHNVSEVELARLVQLNALNCIISYRFGHGVLVKTKDLDLDEQPNVILPI
GPEYLDYVRDLGVDFDHTTVKVNKNNSLDVAKIHKRLIYENFDYILKRVVP
CVTESLLDILPEKTVIERIENFLFYKDESLVOLQDALSATTSLSALTQSS
NNQGGSLSSFLRKQNSNNHSDISNLNLDLSOELARKSNLNNMGFMFTATLS
ASLEVIKDYLDREALIKAKIGADTKETPLRSNEQAKGLNGKGGDRSNYDFL
KVGQEQIENSCNLKTFYGLDMKFNLSLSEQKVERDIKLIELYSKYNQLIQSS
FDNEELAILKEKLFSP"
1309..2001
/gene="BBN02"
1309..2001
/gene="BBN02"
/note="hypothetical protein; identified by Glimmer2;
putative"
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF07682.1"
/db_xref="GI:6382370"
/translation="MTEKSEKEDLQADKEEQNKADTKVISAOEFEEYMRPEQANS
KSKETSRDLSINERITKELAEVERIKQLLEAERINEIDTLAKAHLNHNKEV
LLAGYTKDIMQORRELYRKFVPIEQKAIKAVSDISHIDGEMLEQLVSLAKVNIK
LRKTASNFSSVDSVRGNIIVKSEERASLSDSNFVPINFTEFVQALISNTYKORRIQFY
ENLKRHKRTGIA"
2013..2579
/gene="BBN03"
2013..2579
/gene="BBN03"
/note="hypothetical protein; identified by Glimmer2;
putative"
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF07683.1"
/db_xref="GI:6382371"
/translation="MSDGIITKKEEPDKKVAIEKALMNKPNQDAGLLRNSVDFRDKNL
IYNSDGVFTSSKDKIENYPAKGYPKRGVLSFSADGTTTELEVEAGGGDLYGICTD
IDFTGATVVPVPIINFTGYLTFFKNGQNGVNPDKLHFNAQGELEKXNGGDKSVNAI
ALSKVHLKLTBELSIVLASVFGNALKGN"
2583..3350
/gene="BBN04"
2583..3350
/gene="BBN04"
/note="hypothetical protein; identified by Glimmer2;
putative"
/transl_table=1
/product="hypothetical protein"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

Db 22517 TTGCA 22522

Search completed: February 16, 2004, 15:27:47
Job time : 3247 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:12:22 ; Search time 20 Seconds
(without alignments)
711.648 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDIDVNLLEEQYELG.....NLVKEVGTQVSENPDMMW 148

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR, 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	28.9	198	2 S63233	hypothetical prote
2	132	17.1	134	2 T41221	hypothetical prote
3	116	15.0	404	2 H84593	hypothetical prote
4	97.5	12.6	610	2 C70126	DNA mismatch repai
5	94	12.2	627	2 E70122	flagellar hook-ax
6	92	11.9	270	2 C82114	probable flagellar
7	91.5	11.9	1005	2 A64465	hypothetical prote
8	91	11.8	239	2 AD0088	probable flagellar
9	91	11.8	2325	2 T15566	hypothetical prote
10	90.5	11.7	746	2 T47237	myosin II heavy ch
11	87.5	11.3	755	2 T34567	hypothetical prote
12	87	11.3	263	2 E70215	hypothetical prote
13	86.5	11.2	212	2 AH2423	phosphoglycolate p
14	86	11.1	2285	2 T12796	probable transglyc
15	85	11.0	507	2 A71622	hypothetical prote
16	85	11.0	805	2 A75014	hypothetical prote
17	85	11.0	1553	2 S67483	adenosinetriphosph
18	84.5	10.9	852	2 D72330	conserved hypothet
19	84.5	10.9	1188	2 E89896	chromosome segrega
20	84	10.9	1269	2 F84730	probable myosin he
21	83.5	10.8	592	2 C70302	conserved hypothet
22	83.5	10.8	720	2 G97229	pullulanase [impor
23	83.5	10.8	879	2 C71083	conserved hypothet
24	83.5	10.8	880	2 F75103	conserved hypothet
25	83.5	10.8	909	2 C97325	hypothetical prote
26	83.5	10.8	1231	2 S70553	chromosome-associ
27	83	10.8	777	2 C85065	kinesin-like prote
28	82.5	10.7	244	2 H71873	hypothetical prote
29	82.5	10.7	981	2 G90601	DNA polymerase III

30	82.5	10.7	1225	2 A49464	chromosome segrega
31	82	10.6	238	2 T19175	hypothetical prote
32	82	10.6	1164	2 T24806	hypothetical prote
33	82	10.6	1516	2 E71619	RAD2 endonuclease
34	82	10.6	1963	1 MWKM	myosin heavy chain
35	81	10.5	320	2 AH0942	6-phosphofructokin
36	81	10.5	500	2 E89852	hypothetical prote
37	81	10.5	521	2 B70182	lysine-tRNA ligase
38	81	10.5	564	2 A60115	M protein precursor
39	81	10.5	797	2 E84642	hypothetical prote
40	81	10.5	1188	2 A71621	protein with 5'-3'
41	81	10.5	2429	1 SJHUA	spectrin alpha cha
42	81	10.5	3394	2 T18501	hypothetical prote
43	80.5	10.4	211	2 S69930	outer surface prot
44	80.5	10.4	263	2 B96957	probable transcrip
45	80.5	10.4	402	2 T23701	hypothetical prote

ALIGNMENTS

RESULT 1

S63233
hypothetical protein YNL260c - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N0838
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C:Accession: S63233
R:Sen-Gupta, M.; Gueldeher, U.; Beinhauer, J.; Piedler, T.; Hegemann, J.H.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63220
A:Accession: S63233
A:Molecule type: DNA
A:Residues: 1-198 <SEN>
A:Cross-references: EMBL:Z71536; NID:gl302312; PID:e239647; PID:gl302313; GSPDB:GN00014;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YNL260c
A:Cross-references: SGD:S0005204
A:Map position: 14L

Query Match 28.9%; Score 223; DB 2; Length 198;
Best Local Similarity 35.9%; Pred. No. 1.4e-10;
Matches 47; Conservative 30; Mismatches 40; Indels 14; Gaps 3;

QY	4	IDIDVNLLEEQYELGFKQIQGTDQVLEGEKGYGTGQFRLIIGYIQELMKFWLS	63
DB	37	MDFDNLLNLEEQYQEGFLEGQENIKQSFLGKQYGLQVGFQFRTLLGOMEGLCDV---	93
QY	64	HIDQYN-NSSLRNHLNLEIDMAQISITNGDKVEDEYKNIKAKRNKLRLVIAISITKETW	122
DB	94	-IESYGLHSPTLEKNIHTIRTLMKLKNNDDESVMFEFVLILKLNKFRTI-----	144
QY	123	KIDSLDNLVKE	133
DB	145	-LITLRLVKD	154

RESULT 2

T41221
hypothetical protein SPCC191.08 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T41221
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1998
A:Reference number: Z21904
A:Accession: T41221
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-134 <LYN>
A:Cross-references: EMBL:AL049644; PIDN:CAB41054.1; GSPDB:GN000066; SPDB:SPCC191.08
A:Experimental source: strain 972h-; cosmid ci91

myosin II heavy chain [imported] - Naegleria fowleri (fragment)
C:Species: Naegleria fowleri
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47237
R:Shaw, D.R.; Sullivan, P.K.; Marciano-Cabral, F.; Ennis, H.L.
submitted to the EMBL Data Library, December 1995
A:Description: Codon usage in Naegleria fowleri.
A:Reference number: Z24413
A:Accession: T47237
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-746 <SHA>
A:Cross-references: EMBL:U43192; PIDN:AB01786.1
A:Experimental source: strain LSE mp; cell type amoeba

Query Match 11.7%; Score 90.5; DB 2; Length 746;
Best Local Similarity 27.7%; Pred. No. 17;
Matches 41; Conservative 28; Mismatches 58; Indels 21; Gaps 5;
QY 5 DIDNVNLEE---EYELGFKGQIGTKDOYLEGKEYGYQTGFQRFLLIGVIGELMKFW 61
Db 147 ELENLDETEGGNQDSQFQLQNELQNERNLQMKSENERLOREL-----EEMKRS 200
QY 62 LSHIDQYNSSSLRNHLNLEDIMAQI-----SITNGKVEDYENIKKARNKLR 112
Db 201 LS-DKQNESTSLDSKVKSLDKIRELTALLERSSKTDLDKRSKMDKEVKRLAQQLQ 258
QY 113 VIASITK-ETWKIDSILNVLKVEGVTQLQ 139
Db 259 ETEQALKGTEQKNDADNRVKQLESELQ 286

RESULT 11
T34567
hypothetical protein DKFP434A128.1 - human
C:Species: Homo sapiens (man)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34567
R:Ottenwälder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A:Reference number: Z21542
A:Accession: T34567
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-755 <OTT>
A:Cross-references: EMBL:AL122120
A:Experimental source: adult testis; clone DKFP434A128
C:Genetics:
A:Note: DKFP434A128.1

Query Match 11.3%; Score 87.5; DB 2; Length 755;
Best Local Similarity 23.7%; Pred. No. 29;
Matches 36; Conservative 39; Mismatches 50; Indels 27; Gaps 8;
QY 3 DIDIDN-VLNLEEEQYELGFKGQIGTKDOYLEGKEYGYQTGFQRFLLIGVIGELMKFW 61
Db 322 DGDIDNCALELARIKQETREKENLVK-EKIKLE-SEIGNTFEKKRISVA-DRKLLKCR 378
QY 62 LSHIDQYNSSSLRNHLNLEDIMAQISITNGD-----KEVEDYENIKKARN 109
Db 379 TAYQDHETSRIQLKGELSLK---ATVNTSSDLEALRNKISIKKDIHEETARLQTKN 435
QY 110 KLRVIAS----ITKETWKID----SLDNLVKE 133
Db 436 HNEIITQKLKEITKETSWEKATNLEDMLKE 467

RESULT 12
E70215
hypothetical protein BBA69 - Lyme disease spirochete plasmid A/lps4
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 28-Jul-2000
C:Accession: E70215

R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Whitson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B. Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
A:Reference number: A70100; WUID:98065943; PMID:9403685
A:Accession: E70215
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-263 <KLE>
A:Cross-references: GB:AE000790; NID:g2690224; PIDN:AAC66287.1; PID:g2690286; TIGR:BBAG6;
A:Experimental source: strain B31
C:Genetics:
A:Genome: plasmid
C:Superfamily: Borrelia burgdorferi hypothetical protein BBI38

Query Match 11.3%; Score 87; DB 2; Length 263;
Best Local Similarity 23.5%; Pred. No. 9.6;
Matches 42; Conservative 32; Mismatches 47; Indels 58; Gaps 8;
QY 7 DNVNLEEEQYELGFKGQIGTKDOYLE---GKE-----YGYQ---TGFORFLIGY 53
Db 55 ENIQNFQKSGDLGASDERPMGTASLXKAIGLEDRKNQYDIQAKITNESNLDIY 114
QY 54 IQE-----LMKFWLSHIDQYNSSSLRNHLNLEDIMAQISITNGKVEDYDK 102
Db 115 IRAVELANENEKMLLXRFLLSLDY-----KKENIETLKEIL-----EKLNNYEN 160
QY 103 NIKKARNKLVIA-----SITKETWKIDSLNVLKVEGVTQLQVSE 142
Db 161 DPKIAANFLYRIALDIQLKLEHLKLSINEKLDLTSKSKED-LEALLFQVKSALQIQE 218

RESULT 13
AH2423
phosphoglycolate phosphatase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AH2423
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguch Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076643.1; PID:gl7134082; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr4944
C:Superfamily: Alcaligenes eutrophus phosphoglycolate phosphatase

Query Match 11.2%; Score 86.5; DB 2; Length 212;
Best Local Similarity 29.5%; Pred. No. 8.2;
Matches 38; Conservative 16; Mismatches 48; Indels 27; Gaps 7;
QY 26 IQGTFKDOYLEGKEYGYQTGF-----ORFLIGVIGELMKFWLSHIDQYNSSSLRN 76
Db 85 IPGIKEALLKELKEDYKLGITNSRNVNTNLSINELDSLDFIYSGVTIFGKTTIINN 144
QY 77 HLNLEDIMAQISITNGKVEDYENIKKARNKLRVIASITKETWKIDSLNVLKVEVG 136
Db 145 VLRO-KQFKQSVIYVGD-ETRDIEAS-KKA--NIKVIA-----VTWGFNSPEILAK---- 191
QY 137 TLQVSENPD 145
Db 192 -----QNPD 195

RESULT 14

T12796
Probable transglycosylase - Bacillus subtilis phage SPBc2
C/Species: Bacillus subtilis phage SPBc2
C/Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999
C/Accession: T12796; A69911
R/Lazarevic, V.; Duesterhoft, A.; Solido, B.; Hilbert, H.; Maue, C.; Karamata, D.
Submitted to the EMBL Data Library, August 1997
A/Description: The complete nucleotide sequence of the Bacillus subtilis SPBc2 prophage
A/Reference number: 217583
A/Accession: T12796
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2285 <LAZ>
A/Cross-references: EMBL:AF020713; NID:G3025478; PID:G3025510; PIDN:AAC13005.1
R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.F.; Eniati, K.D.; Errington, J.; Faret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A/Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
A/Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Ser
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A/Authors: Yoshikawa, H.P.; Zunstein, E.; Yoshikawa, H.; Danchin, A.
A/Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A/Reference number: A69580; MUID:98044033; PMID:9384377
A/Accession: A69911
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-2285 <KUN>
A/Cross-references: GB:Z99115; GB:AL009126; NID:G2634478; PIDN:CAB14053.1; PID:sl183582;
A/Experimental source: strain 168
C/Genetics:
A/Gene: yomI

Query Match 11.1%; Score 86; DB 2; Length 2285;
Best Local Similarity 22.4%; Pred. No. 1.4e+02;
Matches 35; Conservative 34; Mismatches 61; Indels 26; Gaps 6;
QY 5 DIDVNLN-EEQYELGFKGQIGTQDYVLEKGYGYQFGFLLIGY-----IQEL 57
DB 1856 DVDNKNISMEDEDKVYKQIKLQCOQKEAKYIKLEEQKAAKGFPIQEQITEE 1915
QY 58 MKFWLS-----HIDQYNNSSLRNHLNLEIDMAQISITNGDKVEDEYKNIKARN-KL 111
DB 1916 MQNWKQKQKQFNELNLYTKSIKDIYKSLADEVSVI-----YKEMYEKMRDIEL 1964
QY 112 RVIASITKTKWTKIDSLNVLKVEGGTLOVSENPDPM 147
DB 1965 EAHQKATQDL--IDEIDKTDDKAFQKELKQDSI 1998

RESULT 15

A71822
Hypothetical protein PFB0155c - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C/Accession: A71622
R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 282, 1126-1132, 1998
A/Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: A71622
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-507 <GAR>
A/Cross-references: GB:AE001376; GB:AE001362; NID:G3845108; PIDN:AAC71821.1; PID:G384511

A/Experimental source: clone 3D7

C/Genetics:

A/Gene: PFB0155c

Query Match 11.0%; Score 85; DB 2; Length 507;
Best Local Similarity 29.8%; Pred. No. 29;
Matches 28; Conservative 17; Mismatches 31; Indels 18; Gaps 4;
QY 66 DQYNNSSLRNHLNLEIDMAQISIT-----NGDKE---VEDYEKNIKARN 109
DB 53 DEYDNISRSKENINNSKKIKNELSIKDNMHDYIYDDRIYNNDKKNVKSNDKNVKSND 112
QY 110 KLRVIASITKTKWTKIDSLNVLKVEGGTLOVSEN 143
DB 113 K-NVKSNDKNVKSND-KNVKSNDKNVKSND 144
Search completed: February 16, 2004, 14:14:46
Job time : 20 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 09:07:46 ; Search time 2877 Seconds
(without alignments)
6356.136 Million cell updates/sec

Title: US-09-980-054A-11

Perfect score: 447

Sequence: 1 atctcgatatagatataga.....accccgatgatgtgtgtsa 447

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.ov.*

5: gb.pat.*

6: gb.ph.*

7: gb.pl.*

8: gb.pr.*

9: gb.ro.*

10: gb.sts.*

11: gb.sy.*

12: gb.un.*

13: gb.vi.*

14: gb.ba.*

15: em.fun.*

16: em.in.*

17: em.mu.*

18: em.ov.*

19: em.pat.*

20: em.ph.*

21: em.pl.*

22: em.ro.*

23: em.sts.*

24: em.un.*

25: em.vi.*

26: em.htg.hum.*

27: em.htg.inv.*

28: em.htg.other.*

29: em.htg.mus.*

30: em.htg.pln.*

31: em.htg.rod.*

32: em.htg.vrt.*

33: em.htg.hum.*

34: em.htgo.hum.*

35: em.htgo.mus.*

36: em.htgo.other.*

37: em.htgo.mus.*

38: em.htgo.other.*

39: em.htgo.mus.*

40: em.htgo.other.*

41: em.htgo.mus.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	447	100.0	447	6	AX057345	AX057345 Sequence
2	442.2	98.9	447	6	AX240813	AX240813 Sequence
3	442.2	98.9	447	6	AX488762	AX488762 Sequence
4	324.4	72.6	326	6	AX057353	AX057353 Sequence
5	71.4	15.0	253305	3	PFMAL3P7	AL034559 Plasmid
6	70.8	15.8	863	11	CNS065VQ	AL395628 T7 end of
7	70.4	15.7	4865	3	AF364131	AF364131 Anopheles
8	69.2	15.5	67970	3	PFMALIP3	AL031746 Plasmid
9	69.2	15.5	178193	2	AC068949	AC068949 Homo sapi
10	68.8	15.4	174835	2	BX248097	BX248097 Danio rer
11	67.6	15.1	94434	5	AL929250	AL929250 Zebrafish
12	67.2	15.0	170627	2	AC125567	AC125567 Rattus no
13	66.4	14.9	110000	2	PFMAL7P1_04	Continuation (5 of
14	66.4	14.9	286208	2	AC117140	AC117140 Rattus no
15	65.8	14.7	25753	8	AF437291	AF437291 Saccharom
16	65.8	14.7	110000	3	AC116984_3	Continuation (4 of
17	65.6	14.7	27491	9	AC010284	AC010284 Homo sapi
18	65.4	14.6	271546	3	AE014843	AE014843 Plasmid
19	65	14.5	65	6	AX485584	AX485584 Sequence
20	65	14.5	840	8	CNS018QK	AL110675 Botrytis
21	65	14.5	214955	2	BX511081	BX511081 Danio rer
22	65	14.5	254449	3	AE014817	AE014817 Plasmid
23	64.8	14.5	119733	9	HSDJ655C5	AL121956 Human DNA
24	64.8	14.5	250029	3	AE014820	AE014820 Plasmid
25	64.6	14.5	58437	9	BX276094	BX276094 Human DNA
26	64.6	14.5	181845	2	BX005330	BX005330 Danio rer
27	64.6	14.5	347582	3	PFMAL4P1	AL034557 Plasmid
28	64.4	14.4	153751	3	AC116551	AC116551 Dictyoste
29	64.4	14.4	330050	3	PFAL929355	AL929355 Plasmid
30	64.2	14.4	1104	8	SCYML260C	Z71536 S.cerevisia
31	64.2	14.4	2961	8	YSCATX1HF	L35270 Saccharomyc
32	64.2	14.4	33016	8	SCCHX1VL	X96722 S.cerevisia
33	64.2	14.4	182870	3	AC116960	AC116960 Dictyoste
34	64.2	14.4	246611	2	AC111404	AC111404 Rattus no
35	64.2	14.4	348552	3	PFAL929354	AL929354 Plasmid
36	64.2	14.4	349980	6	AX444564	AX444564 Sequence
37	64	14.3	51014	2	AC011856	AC011856 Homo sapi
38	64	14.3	110000	3	PFMALIP2_1	Continuation (2 of
39	64	14.3	152900	5	AL954652	AL954652 Zebrafish
40	64	14.3	303671	1	AE015937	AE015937 Clostridi
41	63.8	14.3	110000	3	AC116984_0	AC116984 Dictyoste
42	63.8	14.3	158035	5	BX005364	BX005364 Zebrafish
43	63.8	14.3	347050	3	PFAL929351	AL929351 Plasmid
44	63.6	14.2	110000	3	AC116957_1	Continuation (2 of
45	63.4	14.2	115866	2	AC135614	AC135614 Pan trogl

ALIGNMENTS

RESULT 1	AX057345	AX057345	Sequence 11 from Patent WO0075305.	447 bp	DNA	linear	PAT 17-JAN-2001
LOCUS	AX057345	Sequence 11 from Patent WO0075305.					
DEFINITION	AX057345	Sequence 11 from Patent WO0075305.					
ACCESSION	AX057345	Sequence 11 from Patent WO0075305.					
VERSION	AX057345.1	GI:12310086					
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							

```
HOECHST MARION ROUSSEL (FR)
FEATURES
  source
    1. 447
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
    1. 447
  CDS
    /note="unnamed protein product"
    /codon_start=1
    /transl_table=12
    /protein_id="CAC22438.1"
    /db_xref="GI:12310087"
    /translation="MSDIDIDNVNLEBEQVELFKGQIQGQYKQVLEKVEYQGTG
    FORLLIIGIELKFWLSDIDQYNNSSLENHNLNLDIVAQISITNGKEVEDYEK
    NIKARNKLRVIAITKITETWKIDSLDLNLVKEVGGLQVSENPDDMW"
BASE COUNT 192 a 38 c 82 g 135 t
ORIGIN
  1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
  2 TTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  3 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  4 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  5 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  6 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  7 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  8 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  9 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  10 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  11 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  12 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  13 AGTGAACACCCCGATGATATGGTGA 447
  14 AGTGAACACCCCGATGATATGGTGA 447
Query Match 100.0%; Score 447; DB 6; Length 447;
Best Local Similarity 100.0%; Pred. No. 1.1e-55;
Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
DB 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
DB 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
DB 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
QY 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
DB 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
QY 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
DB 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
QY 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
DB 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
QY 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
DB 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
QY 421 AGTGAACACCCCGATGATATGGTGA 447
DB 421 AGTGAACACCCCGATGATATGGTGA 447
RESULT 2
AX240813
LOCUS AX240813 447 bp DNA linear PAT 26-SEP-2001
DEFINITION Sequence 51 from Patent WO0160975.
ACCESSION AX240813
VERSION AX240813.1 GI:15797749
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
1 Roemer, T., Jiang, B., Boone, C. and Busey, H.
AUTHORS Roemer, T., Jiang, B., Boone, C. and Busey, H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 51 23-AUG-2001;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
  source
    1. 447
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
BASE COUNT 192 a 38 c 82 g 135 t
ORIGIN
  1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
  2 TTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  3 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  4 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  5 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  6 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  7 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  8 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  9 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  10 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  11 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  12 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  13 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  14 AGTGAACACCCCGATGATATGGTGA 447
  15 AGTGAACACCCCGATGATATGGTGA 447
Query Match 98.9%; Score 442.2; DB 6; Length 447;
Best Local Similarity 99.3%; Pred. No. 5.6e-55;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
DB 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
DB 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
DB 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
QY 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
DB 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
QY 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
DB 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
QY 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
DB 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
QY 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
DB 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
QY 421 AGTGAACACCCCGATGATATGGTGA 447
DB 421 AGTGAACACCCCGATGATATGGTGA 447
RESULT 2
AX240813
LOCUS AX240813 447 bp DNA linear PAT 26-SEP-2001
DEFINITION Sequence 51 from Patent WO0160975.
ACCESSION AX240813
VERSION AX240813.1 GI:15797749
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
1 Roemer, T., Jiang, B., Boone, C. and Busey, H.
AUTHORS Roemer, T., Jiang, B., Boone, C. and Busey, H.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 0160975-A 51 23-AUG-2001;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
  source
    1. 447
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
BASE COUNT 193 a 38 c 80 g 136 t
ORIGIN
  1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
  2 TTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  3 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  4 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  5 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  6 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  7 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  8 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  9 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  10 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  11 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  12 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  13 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  14 AGTGAACACCCCGATGATATGGTGA 447
  15 AGTGAACACCCCGATGATATGGTGA 447
Query Match 98.9%; Score 442.2; DB 6; Length 447;
Best Local Similarity 99.3%; Pred. No. 5.6e-55;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
DB 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
DB 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
DB 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
QY 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
DB 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
QY 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
DB 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
QY 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
DB 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
QY 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
DB 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
QY 421 AGTGAACACCCCGATGATATGGTGA 447
DB 421 AGTGAACACCCCGATGATATGGTGA 447
```

```
HOECHST MARION ROUSSEL (FR)
FEATURES
  source
    1. 447
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
    1. 447
  CDS
    /note="unnamed protein product"
    /codon_start=1
    /transl_table=12
    /protein_id="CAC22438.1"
    /db_xref="GI:12310087"
    /translation="MSDIDIDNVNLEBEQVELFKGQIQGQYKQVLEKVEYQGTG
    FORLLIIGIELKFWLSDIDQYNNSSLENHNLNLDIVAQISITNGKEVEDYEK
    NIKARNKLRVIAITKITETWKIDSLDLNLVKEVGGLQVSENPDDMW"
BASE COUNT 192 a 38 c 82 g 135 t
ORIGIN
  1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
  2 TTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  3 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  4 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  5 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  6 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  7 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  8 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  9 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  10 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  11 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  12 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  13 AGTGAACACCCCGATGATATGGTGA 447
  14 AGTGAACACCCCGATGATATGGTGA 447
Query Match 98.9%; Score 442.2; DB 6; Length 447;
Best Local Similarity 99.3%; Pred. No. 5.6e-55;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
DB 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
DB 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
DB 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
QY 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
DB 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
QY 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
DB 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
QY 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
DB 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
QY 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
DB 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
QY 421 AGTGAACACCCCGATGATATGGTGA 447
DB 421 AGTGAACACCCCGATGATATGGTGA 447
RESULT 3
AX488762
LOCUS AX488762 447 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 6062 from Patent WO02053728.
ACCESSION AX488762
VERSION AX488762.1 GI:22322774
KEYWORDS
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Mitosporic Saccharomycetales; Candida.
REFERENCE
1 Roemer, T., Jiang, B., Boone, C., Busey, H. and Ohlsen, K.I.
AUTHORS Roemer, T., Jiang, B., Boone, C., Busey, H. and Ohlsen, K.I.
TITLE Gene disruption methodologies for drug target discovery
JOURNAL Patent: WO 02053728-A 6062 11-JUL-2002;
Elitra Pharmaceuticals, Inc. (US)
FEATURES
  source
    1. 447
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /db_xref="taxon:5476"
BASE COUNT 193 a 38 c 80 g 136 t
ORIGIN
  1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
  2 TTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  3 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
  4 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  5 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
  6 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  7 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
  8 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  9 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
  10 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  11 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
  12 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  13 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
  14 AGTGAACACCCCGATGATATGGTGA 447
  15 AGTGAACACCCCGATGATATGGTGA 447
Query Match 98.9%; Score 442.2; DB 6; Length 447;
Best Local Similarity 99.3%; Pred. No. 5.6e-55;
Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
DB 1 ATGTCAGATATAGATATAGATAATGTTAAATTTAGAGAGAACCAATATGAATTAGGA 60
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
DB 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
DB 121 TATCAAACTGGGATTTCAACGATTTTAAATCATTTGGTTATATTTCAAGAATTAATGA 180
QY 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
DB 181 TGGTTATCCCATATAGATCAATATATAAATCTCTTCTTCACTTCGGGAATCATTTGA 240
QY 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
DB 241 TTGGAAGATATTATGGCACAAATTTCTATAAGATCGAGATAAAGAGTTCAAGATTAT 300
QY 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
DB 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAA 360
QY 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
DB 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAACTTTACA 420
QY 421 AGTGAACACCCCGATGATATGGTGA 447
DB 421 AGTGAACACCCCGATGATATGGTGA 447
```

Db 1 ATGTCAGATATAGATATAGTATATGTAATTAATTTAGAAAGAACATATGATTAAGGA 60

Qy 61 TTTTAAAGAGGTCAAAATACAGGAACAAAGATCAATATTTTGAAGGAAAGATATGCT 120

Db 61 TTTTAAAGAGGTCAAAATACAGGAACAAAGATCAATATTTTGAAGGAAAGATATGCT 120

Qy 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAGAAATTAATGAATTT 180

Db 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAGAAATTAATGAATTT 180

Qy 181 TGGTTATCCCATATAGATCAATATATACTCTTCTTCACTTCGGAAATCATTTGAATAT 240

Db 181 TGGTTATCCCATATAGATCAATATATACTCTTCTTCACTTCGGAAATCATTTGAATAT 240

Qy 241 TTGAGATATTTATGCGACAAATTTCTATACGAATGGAGATTAAGAGTTGATTAAT 300

Db 241 TTGAGAAATATTTTGGACAAATTTCTATACGAATGGAGATTAAGAGTTGATTAAT 300

Qy 301 GAAAAAATATTTAAAGGCAAGAAATAATTAAGAGTGATAGTATTAACCTAAAGAA 360

Db 301 GAAAAAATATTTAAAGGCAAGAAATAATTAAGAGTGATAGTATTAACCTAAAGAA 360

Qy 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAATTTTACAAGTT 420

Db 361 ACTTGGAAATTTGATTCATTTGGATTAATTTGGTGAAGAGTAGGTGGAATTTTACAAGTT 420

Qy 421 AGTGAATCCCGATGATATGTGTGA 447

Db 421 AGTGAATCCCGATGATATGTGTGA 447

RESULT 4

AX057353

LOCUS AX057353 326 bp DNA linear PAT 17-JAN-2001

DEFINITION Sequence 19 from Patent WO0075305.

ACCESSION AX057353

VERSION AX057353.1 GI:12310094

KEYWORDS

SOURCE Candida albicans

ORGANISM Candida albicans

REFERENCE 1 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.

TITLE Lalanne, J.L. and Rocher, C.

JOURNAL Novel candida albicans genes and proteins coded by said genes

Patent: WO 0075305-A 19 14-DEC-2000;

HOECHST MARION ROUSSEL (FR)

FEATURES

source

1..326

/organism="Candida albicans"

/mol_type="genomic DNA"

/db_xref="taxon:5476"

BASE COUNT 146 a 27 c 53 g 100 t

ORIGIN

Query Match 72.6%; Score 324.4; DB 6; Length 326;

Best Local Similarity 99.7%; Pred. No. 6.6e-39;

Matches 325; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 AGATATAGATAATGTTATTAATTTAGAGAGAACAAATATGAATTAGGATTTAAAGAGG 71

Db 1 AGATATAGATAATGTTATTAATTTAGAGAGAACAAATATGAATTAGGATTTAAAGAGG 60

Qy 72 TCAATACAGGAACAAAGATCAATATTTAGAGGAAAGATATGTTTCAAACTCG 131

Db 61 TCAATACAGGAACAAAGATCAATATTTAGAGGAAAGATATGTTTCAAACTCG 120

Qy 132 ATTTCAACGATTTTATCATTTGGTTATATTCAGAAATTAATGAATTTTGGTTATCCCA 191

Db 121 ATTTCAACGATTTTATCATTTGGTTATATTCAGAAATTAATGAATTTTGGTTATCCCA 180

Qy 192 TATAGATCAATATATTAATCACTCTTCTTCACTTCGGAATCAATTTGGAAGATAT 251

Db 181 TATAGATCAATATATTAATCACTCTTCTTCACTTCGGAATCAATTTGGAAGATAT 240

Qy 252 TATGCGCAAAATTTCTTATAACGAATGAGATTAAGAAGTTGAAGATTTGAAAAAATAT 311

Db 241 TATGCGCAAAATTTCTTATAACGAATGAGATTAAGAAGTTGAAGATTTGAAAAAATAT 300

Qy 312 TAAAAAGCGCAGAAATTAATTAAGAG 337

Db 301 TAAAAAGCGCAGAAATTAATTAAGAG 326

RESULT 5

PFMAL3P7/c

LOCUS PFMAL3P7 253305 bp DNA linear INV 29-JAN-2003

DEFINITION Plasmodium falciiparum MAL3P7, complete sequence.

ACCESSION AL004559

AL004559 AL008974 AL008981 AL008983 AL009015 AL010138

AL010143 AL010154 AL010155 AL010156 AL010160 AL010169

AL010187 AL010189 AL010207 AL010214 AL021885 AL021888

AL034559.4 GI:8052273

VERSION

KEYWORDS HTG; 40S Ribosomal protein S3A; acyl transferase; acylaminoacyl-peptidase; ATP-dependent RNA Helicase; cyclophilin; elongation factor; P49C12.11-like protein; HesB-like domain protein; hisone H2A variant; kinesin-related protein; N-acetylglucosamine-1-phosphate transferase; protein kinase; R-CG7 repeat; R-FA3 repeat; rep11; rep20; rifin; RNA-binding protein; stevor; T-complex protein 1 epsilon subunit; telomere; var.

SOURCE Plasmodium falciiparum 3D7

ORGANISM Plasmodium falciiparum 3D7

REFERENCE 1 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. Bowmen, S., Lawson, D., Basham, D., Brown, D., Chillingworth, T., Churcher, C.M., Craig, A., Davies, R.M., Devlin, K., Feltwell, T., Gentles, S., Gwilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moulé, S., Mungall, K., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Whitehead, S., Woodward, J.R., Newbold, C. and Barrell, B.G.

TITLE The complete nucleotide sequence of chromosome 3 of Plasmodium falciiparum

JOURNAL Nature 400 (6744), 532-538 (1999)

MEDLINE 99376085

PUBMED 10448855

REFERENCE 2

AUTHORS Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D., Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K., Buckee, C.O., Burrows, C., Cherevach, I., Chillingworth, C., Chillingworth, T., Christodoulou, Z., Clark, L., Clark, R., Corton, C., Cronin, A., Davies, R., Davis, P., Dear, P., Dearden, P., Doggett, J., Feltwell, T., Gobler, A., Goodhead, I., Gwilliam, R., Hamlin, N., Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S., Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D., Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N., Lawson, D., Lennard, N., Line, A., Maddison, M., McLean, J., Mooney, P., Moulé, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A., Rabinowitz, E., Rajandream, M.A., Rutter, S., Rutherford, K.M., Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R., Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L., Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and Barrell, B.G.

TITLE Sequence of plasmodium falciiparum chromosomes 1, 3-9 and 13

JOURNAL Nature 419 (6906), 527-531 (2002)

MEDLINE 22355708

PUBMED 12368867

REFERENCE 3 (bases 1 to 253305)

AUTHORS Lawson, D., Bowman, S. and Barrell, B.

TITLE Direct Submission

JOURNAL Submitted (17-DEC-1998) P.falciiparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

COMMENT On or before May 14, 2001 this sequence version replaced gi:2982540, gi:2982541, gi:2982544, gi:2894453, gi:2894464,


```

TITLE      Direct Submission
JOURNAL    Submitted (27-MAR-2001) Biological Sciences, Vanderbilt University,
           390 24th Ave. S., Nashville, TN 37235, USA
FEATURES   Location/Qualifiers
source     1..4865
           /organism="Anopheles gambiae"
           /mol_type="genomic DNA"
           /db_xref="taxon:7165"
mRNA       join(455..670,768..851,948..1372,1571..1670,1753..1854,
           1963..2070,3429..3530)
           /product="putative odorant receptor Or2"
           join(455..670,768..851,948..1372,1571..1670,1753..1854,
           1963..2070,3429..3530)
           /note="G protein-coupled receptor"
           /codon_start=1
           /product="putative odorant receptor Or2"
           /protein_id="AAL35507.1"
           /db_xref="GI:117105033"
           /translation="MLIECPPIGVNVRVWLFWSVLRRLRFLVGGIPVAVLVNFQ
           FLKLSWGDSMLIINGFTVLYFNLTGTLVNNRKFPTFEGVAAYALLENK
           DDIPVLYVTRGRMLSTSNLWGLFISACFTVPLFVPGRLGVGTIPGVDLAT
           PTYQVFLVQLVITFPACCMYIPPTSFYATCTLFALVOIALKQBLRGLRHSGTMA
           TGHSAQTLFAELKECKIKHQIQIVHDLSNLVTHLCLEFLSFGNMLCALFLFLSIS
           NLAQIMTGSYIFMLISQMFAYWEANVLEISGIGDAIYNGAWDFEPIRKLII
           LIIAQRPNVIVKGVNYPMTLEMFQKLLNVSYFTLLRVYN"
BASE COUNT 1297 a 1024 c 954 g 1590 t
ORIGIN
Query Match      15.7%; Score 70.4; DB 3; Length 4865;
Best Local Similarity 51.0%; Pred. No. 0.18; Indels 7; Gaps 1;
Matches 196; Conservative 0; Mismatches 181;
QY      4 TCAGATATAGATAGATAGATATGATTAAATTTAGAGAGAACATATGAAATGAGATT 63
DB      2819 TAATATTAAAAATAATATTATTAATAATAATAATAATAATAATAATAATAA 2760
QY      64 AAGAAGGTCAATACAGAGCAAGAATCAATATTAGAGAGAAAGATATGTTAT 123
DB      2759 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2700
QY      124 CAACACTGGATTCAACGATTTTAAATCATTTGGTTATTTCAAGAAATTAATGAAATTTGG 183
DB      2699 AATAATGTAATAATAATGAATAACATA-----ATAATAATAATAATAATAATAAT 2647
QY      184 TTATCCCATATAGATCAATATAATACTCTTCTTCACTTCGGAAATCATTTGAATAATTTG 243
DB      2646 TAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2587
QY      244 GAAGATATTATGGCACAAATTTCTATACGATCGAGATGAAGAGTTGAAGATTATGAA 303
DB      2586 AATAATAATAATAATCATATTATTAATAATAATAATAATAATAATAATAATAATAAT 2527
QY      304 AAAAATATTAAAAAGGCAAGAAATAAATTAAAGTGTAGTACTATATTAATAAGAAACT 363
DB      2526 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2467
QY      364 TGGAAATTTGTTTCAATTTGGATAAT 387
DB      2466 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2443

RESULT 8
PFMAL1P3 67970 bp DNA linear INV 29-JAN-2003
LOCUS     Plasmodium falciparum DNA from MAL1P3.
DEFINITION AL031746 AL844501
ACCESSION AL031746.9 GI:6594243
VERSION   AL031746.9 GI:6594243
KEYWORDS
SOURCE    Plasmodium falciparum 3D7
ORGANISM Plasmodium falciparum 3D7
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS  Hall, N., Pain, A., Berriman, M., Churcher, C., Harris, B., Harris, D.,

```

```

Mungall, K., Bowman, S., Atkin, R., Baker, S., Barron, A., Brooks, K.,
Buckee, C.O., Burrows, C., Charvach, I., Chillingworth, C., Corton, C.,
Chillingworth, I., Christodoulou, Z., Clark, L., Clark, R., Corton, C.,
Cronin, A., Davies, K., Davis, P., Dear, P., Dearden, F., Doggett, J.,
Feltwell, T., Goble, A., Goodhead, I., Gwilliam, R., Hamlin, N.,
Hance, Z., Harper, D., Hauser, H., Hornsby, T., Holroyd, S.,
Horrocks, P., Humphray, S., Jagels, K., James, K.D., Johnson, D.,
Kerhornou, A., Knights, A., Konfortov, B., Kyes, S., Larke, N.,
Lawson, D., Lennard, N., Line, A., Maddison, M., Mclean, J., Mooney, P.,
Moule, S., Murphy, L., Oliver, K., Ormond, D., Price, C., Quail, M.A.,
Rabbinowitsch, E., Rajandream, M.A., Rutter, S., Rutherford, K.M.,
Sanders, M., Simmonds, M., Seeger, K., Sharp, S., Smith, R., Squares, R.,
Squares, S., Stevens, K., Taylor, K., Tivey, A., Unwin, L.,
Whitehead, S., Woodward, J., Sulston, J.E., Craig, A., Newbold, C. and
Barrell, B.G.
Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13
Nature 419 (5906), 527-531 (2002)
22255708
12368867
2
Murphy, L., Pain, A., Berriman, M., Bowman, S., Churcher, C., Harris, B.,
Harris, D., Lawson, D., Quail, M., Rajandream, M., Hall, N. and
Barrell, B.
Direct Submission
Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium,
The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA, UK
On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project,
see http://www.sanger.ac.uk/projects/P_falciparum.
Location/Qualifiers
1..67970
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
/isolate="3D7"
/db_xref="taxon:36329"
/chromosome="1"
complement(join(201..257,455..586,748..804))
/gene="PFA0575c"
note="synonym: MAL1P3.01b"
complement(join(201..257,455..586,748..804))
/gene="PFA0575c"
note="synonym: MAL1P3.01"
complement(join(201..257,455..586,748..804))
/gene="PFA0575c"
note="synonym: MAL1P3.01b"
protein_id="CAD49075.1"
/db_xref="GI:23477009"
/translation="MENRSKTIISQNTIKAHVEANDECKEKEKYLKCFNNYKNFLK
GDLTQACDDYEDYQICVLNDLNKKGLHLSNVSEK"
complement(join(1392..1461,1752..2598,2748..2848,
2990..3276))
/gene="PFA0580c"
note="synonym: MAL1P3.01"
complement(join(1392..1461,1752..2598,2748..2848,
2990..3276))
/gene="PFA0580c"
note="Originally annotated as Plasmodium falciparum;
conserved hypothetical protein but pfam match and other
evidences suggest that it is a putative deoxyribonuclease;
earlier start site possible to make a 412 aa protein
pfam match to entry PF01026 TatD DNase, TatD related
DNase, score 117.30, E-value 2.9e-31
Similar to Plasmodium falciparum conserved hypothetical
protein, upf0006 family malip3.01 SWALL:Q5U0N6
(EMBL:AL031746) (412 aa) fasta scores: E(): 3.7e-150, 100%
id in 412 aa, and to Saccharomyces cerevisiae putative

```

```

deoxyribonuclease yb1055c yb1055c or yb10512 or yb10511
SWALL:YB5 YEAST (SWALL:P34220) (418 aa) fasta scores:
E(): 2.5e-19, 30.47% id in 397 aa"
/codon_start=1
/product="Tad-like deoxyribonuclease, putative"
/protein_id="CAD49076.1"
/db_xref="GI:23477010"
/translation="MKLVPHYIKYINLVFVIGIILKNSLKIYNDLAVISTVNVKYV
LQIKGNSJKNHNRKMDNNESSFDIGSNTLTDQMGVYNSKKGHELDQNLNRAK
NNVDKILITCTCLAEIDSLKICETYPDEGLSAGVHPTNCYEFIDKNKHEER
ITAKGEYEFIKYRNEQVNSKMGNGKKICDGEKMMNNEILLENLDTIPGKY
NEKDKEYLENLKNKIKYIPNRCVICIGELDFRLYFCSKYIQIFYQLKVGQFN
LPLFLMRNSNTEFFKVIDYKPLFKNGGVHSFTDKEDIVHIIVQVKNLYIGVNG
CSLKSNTNNAVKKIPLNILLTDPAGVGVKTHASVEYIKDTVEKRAYTLNKKIK
IHKDDNTYFKERNEFYNIADIAITYKYREAVPDLFCKK"
5005..5496
/genes="MALIP3.02"
/notes="synonym: PFA0585w"
5005..5496
/genes="MALIP3.02"
/notes="contains possible signal sequence. Signal peptide
predicted SignalP 2.0 HMM (Signal peptide probability
0.840, signal anchor probability 0.357) with cleavage site
probability 0.504 between residues 27 and 28. ScanRegExp
hit to P800867, Carbamoyl-phosphate synthase subdomain
signature 2"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/db_xref="SP:PREMEL:Q9U0N5"
/translation="MKLLNRRFVLCPIILFFFLNSVVLGNRRNNINHFHETENAAK
AMKLLSGEINSLKLDNGDELKLDNKHDKDSTKWDKSYSPISLEEKYSQTLDFR
KQGEINENTKILEDQEPYILNDSIENIATRFVLENNFDELYIQSFQSLIDIIQS
LN"
8179..10303
/notes="Putative centromere"
14884..20352
/genes="MALIP3.03"
/notes="synonym: PFA0590w"
14884..20352
/genes="MALIP3.03"
/notes="Putative ABC transporter
pfam match to entry PF00005 ABC tran, ABC transporter,
score 45.00, E-value 1.7e-09; HMMSmart hit to SM00382,
Atfases associated with a variety of cellular activities;
ScanRegExp hit to P800211, ABC transporters family
signature
Similar to Schizosaccharomyces pombe ATP-binding cassette
transporter abc1 abc1 or spacs9.12C SWALL:ABC1_SCHPO
(SWALL:Q92337) (1427 aa) fasta scores: E(): 4.2e-08,
21.46% id in 1146 aa"
/codon_start=1
/product="ABC transporter, putative"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/db_xref="GOA:Q9U0N4"
/db_xref="SP:PREMEL:Q9U0N4"
/translation="MTYIKGVNIGSNKGNKKKSCQNI SPLNLSFDWIRPLINDLIK
GDIDELPNI CRNFDPVYASGLENDLIEVDESEFYSEKNSNEVHHCNDSASE
KQVNVYVYHNLWILKTFPRILIIISFYLETILVTLGGKFDYMRILEGQKIPV
YISLPDKPVFSGVVMVIMFFLIPFELALHFYFLFTINLKVSLMYFLYKLNCSN
NHLQNPDAFYNTYKFSQSQTEIDISIRDFLSIGKNASSSSSGIKNNKXNIDNNKFVN
DYITNFIKSTKMKESKSLNRSPLPNVNIYIMFSDVPSVTFPVTCINLNVFVKIF
MSFVFIHKGSNEVGTALWLSALYSAMILFPLSLFKSKYLIYRDKLDMNHVL
KEFLIKMPFNSFAFKIINIFRMKENKYKIRLYISNIGVFISISSDI VEVVIFFI
YKDLRLNKEIEKETSIIIMPLYIKILISVANFPNLVNNMVEGIVNKLKLNININH
LYNDIKNYRMYRTYRIVVDKTLQNDITISHDGTSNHLKHLKLVNKKUTN
MKFYFFPYHKNYTHKNIINQILSLGLNVDNTNKKICFQEHKSNSTYNYSNHHIE
KKEVEYTHNSNSTMNEPFEKKNNEYI IKLENCISFLGSDNKCNDHILKNINFN
LKRSALIIIGNVSGKSARFHSILGDFNMTHGNYIENFFKMPILYVFPQNSWLFPG
NIRSMILPGNEVNLVYKVTILOSELNDLSTIEHGDMPKYNIDNHLNLSKQVRICLA
RALYEHYHMKLCTDEKLLIQNEILDKDLINNKISSNNKSKLVNINYPFNE

```

```

YLOCLMDNNFVLYLLDDIFTSLDPSISKIPSNLFCCKEDNISFKDNCFSFIISMNKS
TLDNFLIEDLVNQVENVNI FEIQDKTLKYEGNISYEMKNNLNITKESHWGYSLNLT
IDYTRILKLFDEVLNHHVGNKMIYKEAYFYKGNTEVSVEFISINSKEYIKKKKKNY
KKEHMKNNNNNNNNNNKDDHINIMNDNHRNYNDINGENSTDDSDPVSIGNE
YTLDTYTSNNSDEEIVKPLYKDTHEFNKSSMPFVKSSNNMINNPNSPKYEINSSS
FKGISILETYLVFPQGVFLVTSVIFMLISITDEIKFVFLTMMISIIISNNKEHSD
TILQKVRYLEYVILPFIISLVTSGICFMSIYGNITSALKVHNHIIISLNAPLYIF
YNNLGNINRFIIDISAPDVGPKRIYKAPFIFFRCILSSLLIIMRDCIFIFPVP
IILIFYFVKPSRSGCEAORLVLSCHTCLNLYSNALSGKNIINIKKNTYHLDVVE
HYNNPRISYFFKWLINIMASLYIKIFILLITTYIMPHLYASGIIKLYKKNYVRI
LSTIGYICISARLVGVIILFELCDYTHIEKEMCCVQRLERFAKISKENASAKENEL
NVITQYCKENENISDKISAIYKVNLUSSINSODDESKKYGKIKENVTVSYK
KXIPLVNGTYKIDDEEPLNINMYALKNOKIGIVGSGAGKSTILLIILINISOG
KITVEGRDITRYNKGEDSIIGILAOISVFNWNI RTFDVYNFTDDIIVHAKLN
GINKNDILYKMHKDNKSNYKIIQTSKVINOSNDNTIITNDICRYLSLVALYN
RHKYIILIDIEPIFNLNNSVHDELNLGKAKSPNYIIRNHPNNTVLIISHANT
LSCDDIYVLRKGEITFYRCSYEDVKTSLSHLEMD"
23896..26044
/product="18S ribosomal RNA A-type"
/notes="18S rRNA expressed in asexual stage parasites"
26045..26438
/product="ITS1 A-type"
/notes="ribosomal operon internal transcribed spacer 1
expressed in asexual stage parasites"
26439..26559
/genes="mal1 5.8S rRNA"
26439..26559
/genes="mal1 5.8S rRNA"
/product="5.8S ribosomal RNA"
/notes="5.8S ribosomal expressed in asexual stage
parasites"
26560..27435
/product="ITS2 A-type"
/notes="ribosomal operon internal transcribed spacer 2
expressed in asexual stage parasites"
27436..31539
/product="28S ribosomal RNA (A-type)"
/notes="28S rRNA expressed during asexual development"
complement(join(31966..32476,32675..32775))
/genes="MALIP3.04"
/notes="synonym: PFA0610c"
complement(join(31966..32476,32675..32775))
/genes="MALIP3.04"

Query Match 15.5%; Score 69.2; DB 3; Length 67970;
Best Local Similarity 50.5%; Pred. No. 0.13; Indels 5; Gaps 2;
Matches 222; Conservative 0; Mismatches 213;

QY 6 AGATATAGATATAGATATATGTTAAATTTAGAGAAGAACAAATATGAAATTTAA 65
Db 8476 ATATAATTTAAATAATAAATACTATTATTAATAATAATTAATAATAATAA 8535

QY 66 AGA--AGGTCAATCAAGGACAAAGACATCAATTTTAGAGAGGAAGAATATGTTAT 123
Db 8536 ATATTATATTAAATAATAATAATAATAATAATAATAATAATAATAATAATA 8595

QY 124 CAACCTGGATTTCAACGATTTTAAATCATTCGTTATATTTCAAGAAATTAATGAAATTTTGG 183
Db 8596 TAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8655

QY 184 TTATCCCATATAGATCAATATAAATCACTCTCTTCCTTCACITCGGAATCATTTG---AATTAAT 240
Db 8656 ATTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8715

QY 241 TTGGAGATATTATATGCGCAAAATTTCTATACGAATGGAGATAAAGAGATTGGAAGATTAT 300
Db 8716 ATATTATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 8775

QY 301 GAAAAAATAATTAATAAGCAAGAAATAATAATAAGAGTGATAGTAGTAACTAAGAA 360
Db 8776 TAAATAATTAATAATAAGAAATAATAATAATAATAATAATAATAATAATAATAATA 8835

```

Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

[illegible]


```

RESULT 13
PFMAL7P1_04/C
WFCOMMENT
Sequence split into 14 fragments LOCUS PFMAL7P1 Accession AL844506
      Fragment Name      Begin      End
PFMAL7P1_00             1          110000
PFMAL7P1_01             1          210000
PFMAL7P1_02             1          310000
PFMAL7P1_03             1          410000
PFMAL7P1_04             1          510000
PFMAL7P1_05             1          610000
PFMAL7P1_06             1          710000
PFMAL7P1_07             1          810000
PFMAL7P1_08             1          910000
PFMAL7P1_09             1          1010000
PFMAL7P1_10             1          1110000
PFMAL7P1_11             1          1210000
PFMAL7P1_12             1          1310000
PFMAL7P1_13             1          1410000

```

Cleveland, C.: Cockrell, R.: Cox, C.: Coyle, M.:

DEMAY
E 11717 7

Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, I., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvaty, Y., Havlak, P., Haves, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A., Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschewa, L., Loulseg, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelenah, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Popper, F., Poudexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sleson, I., Sitter, C.D., Smajd, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimele, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, C., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niedenhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Direct Submission
 2 (bases 1 to 286208)
 Worley, K.C.
 Direct Submission
 Submitted (08-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 286208)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 6, 2002 this sequence version replaced gi:23097379.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information

Center project name: GUCU

Center clone name: CH230-365F7

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 125022 bases at least Q40

Consensus quality: 128107 bases at least Q30

Consensus quality: 129968 bases at least Q20

Estimated insert size: 130783; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: this is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved

* 1 47440: contig of 47440 bp in length

* 47441 47540: gap of unknown length

* 47541 230268: contig of 182728 bp in length

* 230269 230368: gap of unknown length

* 230369 254735: contig of 24367 bp in length

* 254736 254835: gap of unknown length

* 254836 275940: contig of 21105 bp in length

* 275941 276040: gap of unknown length

* 276041 277252: contig of 1212 bp in length

* 277253 277353: gap of unknown length

* 277354 278494: contig of 1141 bp in length

* 278495 282220: gap of unknown length

* 282221 283220: contig of 3627 bp in length

* 283221 284049: contig of 1729 bp in length

* 284050 284149: gap of unknown length

* 284150 286208: contig of 2059 bp in length.

FEATURES

source

1. 286208

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-365F7"

6333..7149

/note="clone boundary"

clone_end:77

site:

end sequence:B2116846"

6368..7302

/note="clone boundary"

clone_end:Sp6

site:

end sequence:B2116847"

47541..48868

/note="wgs end extension"

clone_end:Sp6"

230383..232081

/note="wgs end extension"

clone_end:Sp6"

BASE COUNT 43136 a 24292 c 24002 g 39383 t 155395 others

ORIGIN

Query Match 14.9%; Score 66.4; DB 2; Length 286208;

Best Local Similarity 48.6%; Pred. No. 0.23;

Matches 207; Conservative 0; Mismatches 218; Indels 1; Gaps 1;

Qy 6 AGATATAGATAGATAATGTTATTAATTTAGAGAGAACATATGATTAGATTAA 65

Db 229766 AAATTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 229707

Qy 66 AGAGGTCGAATACAGGACAAACATCAATTTTGAAGGAAAGAAATATGGTTATCA 125

Db 229706 TAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 229647

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: February 16, 2004, 14:28:37 ; Search time 65 Seconds
(without alignments)
1004.995 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDIDVNLNLEBEQVELG.....NLVKEVGGLQVSENDDWV 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Maximum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09980054/runat_13022004_163841_26344/app_query.fasta_1.327
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human-0.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09980054 @Cgn1_1_56 @runat_13022004_163841_26344 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSEBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	91.5	11.9	1664976	4	US-08-916-421B-1
2	86	11.1	7100	4	US-09-308-375-1
3	84	10.9	1664976	4	US-08-916-421B-1
4	83.5	10.8	3258	2	US-08-729-955A-2
5	82	10.6	2271	4	US-09-134-001C-716
6	81	10.5	1563	4	US-09-508-370A-11
7	80	10.4	6773	4	US-09-166-350-27
8	80	10.4	1230025	4	US-09-198-452A-1
9	78	10.1	417	4	US-09-252-991A-6512
10	78	10.1	747	4	US-09-252-991A-6539
11	78	10.1	1521	4	US-08-252-991A-6495
12	78	10.1	3312	3	US-08-923-992A-3

13	77.5	10.0	1488	3	US-08-867-611-9	Sequence 9, Appli
14	77.5	10.0	1488	5	PCT-US92-06965A-14	Sequence 14, Appli
15	77.5	10.0	1497	1	US-08-500-222-1	Sequence 1, Appli
16	77.5	10.0	1497	1	US-08-500-125-1	Sequence 1, Appli
17	77.5	10.0	1497	2	US-07-779-704B-1	Sequence 1, Appli
18	77	10.0	1275	3	US-08-867-611-21	Sequence 21, Appli
19	77	10.0	1275	5	PCT-US92-06965A-26	Sequence 26, Appli
20	77	10.0	1494	4	US-09-328-352-3160	Sequence 3160, Ap
21	77	10.0	1590	4	US-08-887-534A-73	Sequence 73, Appli
22	77	10.0	1590	4	US-09-527-431-73	Sequence 73, Appli
23	77	10.0	1851	3	US-08-867-611-29	Sequence 29, Appli
24	77	10.0	1851	5	PCT-US92-06965A-34	Sequence 34, Appli
25	76.5	9.9	1491	3	US-08-926-842B-35	Sequence 35, Appli
26	76.5	9.9	2433	4	US-09-540-824-24	Sequence 24, Appli
27	76.5	9.9	10917	3	US-08-926-842B-11	Sequence 11, Appli
28	76	9.8	2617	1	US-08-430-024-1	Sequence 1, Appli
29	76	9.8	2617	1	US-08-782-009-1	Sequence 1, Appli
30	76	9.8	2617	3	US-09-017-302-1	Sequence 1, Appli
31	76	9.8	3730	1	US-08-242-932-8	Sequence 8, Appli
32	76	9.8	3730	1	US-08-714-481-8	Sequence 8, Appli
33	76	9.8	3730	5	PCT-US95-06111-8	Sequence 8, Appli
34	76	9.8	4200	1	US-08-242-932-1	Sequence 1, Appli
35	76	9.8	4200	1	US-08-714-481-1	Sequence 1, Appli
36	76	9.8	4200	3	US-08-923-992A-1	Sequence 1, Appli
37	76	9.8	4200	5	PCT-US95-06111-1	Sequence 1, Appli
38	75.5	9.8	777	4	US-09-107-532A-835	Sequence 835, App
39	75.5	9.8	993	4	US-09-134-001C-974	Sequence 974, App
40	75.5	9.8	2043	4	US-09-601-198-181	Sequence 181, App
41	75.5	9.8	2404	1	US-08-484-105-7	Sequence 7, Appli
42	75.5	9.8	2404	1	US-08-484-106-7	Sequence 7, Appli
43	75.5	9.8	3636	4	US-09-134-001C-1983	Sequence 1983, Ap
44	75	9.7	3033	4	US-09-134-001C-2341	Sequence 2341, Ap
45	74.5	9.7	789	1	US-08-160-670A-31	Sequence 31, Appli

ALIGNMENTS

RESULT 1
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
Patent No. 6503729
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g

```

; NAME/KEY: misc feature
; LOCATION: (98120)..(98120)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

Alignment Scores: 451 1664976
Pred. No.: 91.50 Length: 43
Score: Matches:


```

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g

```

```

; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1

```

```

Alignment Scores:
Pred. No.: 4,07e+03
Score: 84.00
Percent Similarity: 42.50%
Best Local Similarity: 26.67%
Query Match: 10.88%
DB: 4
Length: 1664976
Matches: 32
Conservative: 19
Mismatch: 45
Indels: 24
Gaps: 4

```

US-09-980-054A-12 (1-148) x US-08-916-421B-1 (1-1664976)

QY 7 AspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPhe----- 21
DB 1122980 GATAAAATTTGAGTTAATAAGATACGATGAGAAAGGTTCATATGTAGTCTTAAA 1122921

QY 22 LysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyr 41
DB 1122920 AAAGCTGGAAATCAACCTTACAAATCTACATAATGTGACAAATATATCTCTTAAA 1122861

QY 42 GlnThrGlyPheGln-----ArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLys 59
DB 1122860 GAACTGGGATATATTTAAACAACTTTGATTTGGGTGAGATTATGGAAGTTTACCC 1122801

QY 60 PheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsn 79
DB 1122800 TTT-----AAAATCAATCCCTTATTGATCATGTTAAC 1122768

QY 80 AsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAsp 99
DB 1122767 GATATGGTCAATATTTGGGAGAGATAAATATAGA----- 1122732

QY 100 TyrGluLysAsnIleLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLys 119
DB 1122731 TACTTAAAACTATAAAGAGAGCTTTAGAGGCTTTAAACATAAAGTTAGATATCGAAAAA 1122672

RESULT 4

US-08-729-955A-2
; Sequence 2, Application US/08729955A
; Patent No. 5932417
; GENERAL INFORMATION:
; APPLICANT: Birnbaumer, Lutz
; APPLICANT: Zhu, Xi
; TITLE OF INVENTION: Method And Compounds For Controlling
; TITLE OF INVENTION: Capacitative Calcium Ion Entry Into Mammalian Cells
; TITLE OF INVENTION: Essential for Agonist-Activated Capacitative Ca2+
; TITLE OF INVENTION: Entry
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pons, Smith, Lande & Rose
; STREET: 2029 Century Park East, Suite 3800
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: October 15, 1996
; APPLICATION NUMBER: US/08/729,955A
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025,111
; FILING DATE: August 29, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Oldenkamp, David J.
; REGISTRATION NUMBER: 29,421
; REFERENCE/DOCKET NUMBER: 120186
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (310) 788-5000
; TELEFAX: (310) 277-1297

INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3258 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO

; ORIGINAL SOURCE:
; ORGANISM:
; INDIVIDUAL ISOLATE: Mtrp4
US-08-729-955A-2

Alignment Scores: 0.876 Length: 3258
Pred. No.: 83.50 Matches: 37
Score: 45.21% Conservative: 29
Percent Similarity: 25.34% Mismatches: 49
Best Local Similarity: 10.82% Indels: 31
Query Match: 2 Gaps: 6
DB: 2

US-09-980-054A-12 (1-148) x US-08-729-955A-2 (1-3258)

QY 19 LeuGlyPheLysGluGlyGln-----IleGlnGlyThrLysAspGlnTyrLeuGlu 35
DB 2619 TTGAATTTTAAAGAAATGATGTGTGAGCTCATCCAGGGTCAAAAGCAAGCTTCCCAAGAA 2678

QY 36 GlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGln 55
DB 2679 GATGCGAGATGAACAAGAGAAATGAAGAAATTTTGGATTTTCAGAAAGTCCAGAA 2738

QY 56 GluLeuMetLysPhe-----TrpLeuSerHisIleAspGlnTyrAsnAsn 70
DB 2739 GACCTTTCAAAATTTTCACTTGACAAAATTCAGTTGGCACACAAACAATCAACTACA 2798

QY 71 SerSerSerLeuArgAsnHisLeuAsnLeuGluAspIleMetAlaGlnIleSerIle 90
DB 2799 AGGAGCTCAGAGATTATCATTTAAAT-----AGTTTC 2831

QY 91 ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys 110
DB 2832 AGTAACCTTCCAAGACAA-----TATCAGAAATCATGAGAGACTCATTAAGAA 2882

QY 111 LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu 130
DB 2883 TATGTATTGCGGCCCGCCAGATTGATAGAGAGCGATGAGTGAAT-----GAAGGGGAA 2936

QY 131 ValLysGluValGlyGlyThrLeu-----GlnValSer 141
DB 2937 TTGAAGAAATTAAGCAAGACATCTCAAGTCCGTTATGAACCTTGAAGAGAAATCA 2996

QY 142 GluAsnProAspAspMet 147
DB 2997 CAGAACTCAGAGACCTA 3014

RESULT 5

US-09-134-001C-716
; Sequence 716, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 716
; LENGTH: 2271
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-716

Alignment Scores: 0.828 Length: 2271
Pred. No.: 82.00 Matches: 35
Score: 43.42% Conservative: 31
Percent Similarity:

```

Best Local Similarity: 23.03% Mismatches: 58
Query Match: 10.62% Indels: 28
DB: 4 Gaps: 6

US-09-980-054A-12 (1-148) x US-09-134-001C-716 (1-2271)

Qy 7 AspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIle 26
Db 649 GAAGATGTGATTGGTTTACGTGAGGAATTATCAGAA-----CAATAT 690
Qy 27 GlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyGln----- 42
Db 691 CGCTCACTTCAAGATTAAAGAAATTAGCAGCAAAATRTGGATTGATATTAGCGCTCT 750
Qy 43 ---ThrGlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrp 61
Db 751 GCTACTAACTCAAGAA-----GCTGTGCAATGGTTATCTTACCATAT 795
Qy 62 LeuSerHisLysAspGlnTyrAsnAsnSerSerLeuArgAsnHisLeuAsnLeu 81
Db 796 TTAGCTGCTATCAAGACAAACAAATGGTCGAGCAATGAGTTAGGAGCTACTTCAACATTC 855
Qy 82 GluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGlu 101
Db 856 TTAGATATTATGCTGAACGTGATTTACAAATGGTGACATCACTGAACAGAAAGTTCAA 915
Qy 102 LysAsnIleLysLysAlaArgAsnLysLeuArgValIleAla---SerIleThrLysGlu 120
Db 916 GAAATCATTGACCACCTTCATTATGAAATTACGTATCGTTAAATTCGCACGTACGCTGAA 975
Qy 121 -----ThrTrpLysIleAspSerLeuAspSerLeuAsnLeuVal 131
Db 976 TATAATGAATATTCCTCTGAGATCCCACTTGGTAACCTGAAATCATCGTGGTGTA--- 1032

Qy 132 LysGluValGlyThrLeuGlnValSerGluAsn 143
Db 1033 ---GCTATTGACGGCGCTCGATGGTAACTTAAAC 1065

RESULT 6
US-09-508-370A-11
; Sequence 11, Application US/09508370A
; Patent No. 6492131
; GENERAL INFORMATION:
; APPLICANT: Michael Ibba
; APPLICANT: Dieter Soll
; TITLE OF INVENTION: Class I-type Lysyl-tRNA Synthetase
; FILE REFERENCE: OCR-896
; CURRENT APPLICATION NUMBER: US/09/508,370A
; CURRENT FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: PCT/US98/18968
; PRIOR FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MS Dos
; SEQ ID NO 11
; LENGTH: 1563
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: lysyl t-RNA synthetase
; OTHER INFORMATION: clone described in Example 3
US-09-508-370A-11

Alignment Scores:
Pred. No.: 0.663 Length: 1563
Score: 81.00 Matches: 32
Percent Similarity: 36.24% Conservative: 22
Best Local Similarity: 21.48% Mismatches: 49
Query Match: 10.49% Indels: 46
DB: 4 Gaps: 4

US-09-980-054A-12 (1-148) x US-09-508-370A-11 (1-1563)

Qy 14 GluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyr 33

```

```

Dy 1033 AAAAGAGAAAAAAGAGCAATTAA-----AGATTACGAACTATCTCAACCATAC 1086
Qy 34 LeuGluGlyLysGluTyrGlyThrGlyPheGlnArgPheLeuIleGlyTyr 53
Dy 1087 ATCCAGCAAGAAAGAAATCCCTTATCAGTCGAGATCAGACATTAAGTGTAATCAGTCAA 1146
Qy 54 IleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsnSerSer 73
Dy 1147 ATATTTCGAA-----AATAATATAATAATAA 1170
Qy 74 LeuArgAsnHisLeuAsnLeuGluAspIleMetAla----- 86
Dy 1171 ATTTTAATTTACTTGAAAAACGTTCAAGAAAGATCAAAAAGACAAACTATAATAATAA 1230
Qy 87 -----GlnIleSerIleThr 91
Dy 1231 AATTGGCAATTAATTGGATAAGAGATTTTGACCCGAGAGATTTCAAAATTTTCATTAAGA 1290
Qy 92 AsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeu 111
Dy 1291 TCTAAATTTGATATATGGAATATCTAGAGAAATAGCAAAAAGCAATTAATGACTT 1350
Qy 112 -----ArgValIleAlaSerIleThrLysGluThr 121
Dy 1351 TTGGATTTTTRAAAGAAAAAATTTGAAGTTGCCACAGACAAAGACATTCAAAACCGAAATA 1410
Qy 122 TrpLysIleAspSerLeuAspAsnLeu 130
Dy 1411 TATAAATTTCAAGAGAAATATATATA 1437

RESULT 7
US-09-166-350-27
; Sequence 27, Application US/09166350A
; Patent No. 6440663
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Matthew
; APPLICANT: Chen, Yao
; APPLICANT: Stockert, Elisabeth
; APPLICANT: Old, Lloyd
; APPLICANT: Jager, Eike
; APPLICANT: Knuth, Alex
; TITLE OF INVENTION: Renal Cancer Associated Antigens and
; FILE REFERENCE: L0461/7051
; CURRENT APPLICATION NUMBER: US/09/166,350A
; CURRENT FILING DATE: 1998-10-05
; EARLIER APPLICATION NUMBER: US 09/166,350
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 6773
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-166-350-27

Alignment Scores:
Pred. No.: 6.71 Length: 6773
Score: 80.00 Matches: 35
Percent Similarity: 40.44% Conservative: 20
Best Local Similarity: 25.74% Mismatches: 54
Query Match: 10.36% Indels: 27
DB: 4 Gaps: 4

US-09-980-054A-12 (1-148) x US-09-166-350-27 (1-6773)

Qy 9 ValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGly 28
Dy 3311 GTAAATGAACCTTGAGAACTTCAGGTACAACTTCAAAAGAAAAAGAAACAGCTTCAGAAA 3370
Qy 29 ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThrGlnThrGlyPheGlnArgPhe 48

```

Db 3371 ACCATGCAAGATTAGAGCTGTTAAAAAGGATGCCCAACAAACACCATTTGATGAATATG 3430
QY 49 LeuIlelleGlyTyrIleGlnLeuMetIlySpheTrpLeuSerHisIleAspGlnTyr 68
Db 3431 GAATAGCTGATTAT--GAACGTTTGATGAAA----- 3460
QY 69 AsnAsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle 88
Db 3461 GAACATAATCAAAGCTTAACATAATAAAACAAACAAAGATGAGAGATTTGGACCAAGAAATA 3520
QY 89 -----SerIleThrAsnGlyAspIlySglu 96
Db 3521 AAAATTCAAAACACAGAAACCAAGAACCTTACAGAGAAATAACTTCAATGACTTTCATCA 3580
QY 97 ValGluAspTyrGluIleAsnIleIlySylasalaArgAsnIlySLeuArgValIleAlaSer 116
Db 3581 GTACACAAATATGAGAAAA-----AACACCAAAATCAACCAATTCGTTGTG 3628
QY 117 IleThrIlySglu-ThrTrpIlyleAspSerLeuAspIlySLeuVal 131
Db 3629 AAACCAAAAGGAAGAACTGCAGATTCAAAGCAAGCAAGAACTGATC 3674
RESULT 8
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 655294
; GENERAL INFORMATION:
; APPLICANT: Grifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (885001)..(900000)

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature
LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc feature

Alignment Scores:
Pred. No.: 8.69e+03 Length: 1230025
Score: 80.00 Matches: 39
Percent Similarity: 39.86% Conservative: 16
Best Local Similarity: 28.26% Mismatches: 47
Query Match: 10.36% Indels: 36
DB: 4 Gaps: 6

US-09-980-054A-12 (1-148) x US-09-198-452A-1 (1-1230025)

QY 7 AspsanValleuAsnLeuGlugluGlnTyrgluLeuGlyPheLysGluGlyGlnIle 26
Db 199813 GACGAAGTCCTTGCAATGAAGATCAATG-----GAAGCTCAGCTT 199854
QY 27 GlnGlyThrLysAspGlnTyr-----LeuGluGlyLysGluTyrglyTyrglnThrGly 44
Db 199855 CAGGATTTAAGGACCTTGTAGTACCTGGGAGGAAAA----- 199893
QY 45 PheGlnArgPheLeuIleGlyTyrlleGlnGluLeuMetLysPheTrpLeuSerHis 64
Db 199894 -----TATCAGGARTTTAAGAAAAACAAGCTTTCCAAA 199926
QY 65 IleAspGlnTyrAsnAsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIle 84
Db 199927 GTTTAGTTTACGACTTCACAAAATCCTATTCTTAACCTCTTAATCGTTTGGAGGTACTC 199986
QY 85 MetAlaGlnIleSer-----IleThrAsnGlyAspLysGluValGluAspTyrr 100
Db 199987 CATGCCGAGAGCTCCACGAGTATTGTGTATTACATGTCTAGATAGATCGGAAGATCTG 200046
QY 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
Db 200047 AAGAAAAACATCGAGGAGATTGACCGCAATT-----TTTCAGTAACCTCCTGAA 200097
QY 121 ThrTrpLysIleAspSerLeuAsnLeuValLysGluValGlyGlyThrIleu 138
Db 200098 -----GAGCTCTCTTTGTAGTCGGGAATATCAGGACTCATG 200136

RESULT 9

US-09-252-991A-6512
; Sequence 6512, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107195.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6512
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6512

Alignment Scores:
Pred. No.: 0.259 Length: 417
Score: 78.00 Matches: 29
Percent Similarity: 38.93% Conservative: 22
Best Local Similarity: 22.14% Mismatches: 34
Query Match: 10.10% Indels: 46
DB: 4 Gaps: 5

Qy	29	ThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgpHe	48
Db	402	TCCCGGACAGGTGTGAGCACTGCATTCGGT-----GGCGAACC CGACATTC	349
Qy	49	LeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyr	68
Db	348	CTCAGTCCCGGCANTGGCGAATCATG-----	319
Qy	69	AsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp	83

Db
697 AACCGCATGCAGACC-----CCGAGGAGACC----- 723

382 ---AATAGACTGTTACACATCAACCAATGAAGATTGAGAAACATAACCAACCTAAC 438
58 -----MetLysPheTrpLeuSerHisIleAspGlnTyTrAsnAsnSerSerSer 73
439 CAACAGAAAACCTCTGAACACAGTCAGATACGAAGTAGATCTAGCAATATTGACAAAGAG 498
74 LeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGly 93
499 CTTAATCATCAAAAAGTCAAGTTGAAGCAATGGCAGAGCAAGCGGGAATCAAAATGAA 558
94 AspLysGlu-----ValGluAspTyTrGluLysAsnIleLysLysAlaArg 108
559 GATAAAGATTCTATGCTGAAAAAATCGAAGATATTCGTAACAACACTCAACAGCGAT 618
109 AsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSer----- 126
619 AAAAAGACAGATCCGAGTAAGTTCTCGTAGAAGTAAGTAGTAAGTAACTCTTAGTTCAACT 678
127 -----LeuAspAsnLeuValLysGlu 133
679 AAAGCTGGTCTGGATCAACAAATTCAGAG 708

RESULT 13
US-08-667-611-9
; Sequence 9, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAMSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; TITLE OF INVENTION: ANTIGENS
; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:

QY 129 AsnLeuValLysGluValIleGlyThrLeuGln 139
Db 724 CGTACATCGTCAGGCGCGGCGACCTCGCAA 756

RESULT 12
US-08-923-992A-3
; Sequence 3, Application US/08923992A
; Patent No. 6280738
; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; APPLICANT: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-Iga Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3312
US-08-923-992A-3

Alignment Scores:
Pred. No.: 4.5 Length: 3312
Score: 78.00 Matches: 36
Percent Similarity: 41.33% Conservative: 26
Best Local Similarity: 24.00% Mismatches: 64
Query Match: 10.10% Indels: 24
DB: 3 Gaps: 5

US-09-980-054a-12 (1-148) x US-08-923-992A-3 (1-3312)

QY 5 AspleAspAsnValLeuAsnLeuGluGluGlnTyTrGluLeuGlyPheLysGluGly 24
Db 268 GATGTGGATAACACATCTATCTCATGACAGAAAATGATTTAAACAAAAATGAT 327
QY 25 GlnIleGlnGlyThrLysAspGlnTyTrLysGluGlyThrGlyTrpGlnThrGly 44
Db 328 GAAACAATGATTCCT--GATGCATTATTAGAAATCAATTAACGAAACT--- 381
QY 45 PheGlnArgPheLeuIleGlyTyTrIleGlnGluLeu----- 57

APPLICATION NUMBER: US 07/448,566
 FILING DATE: 21-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834.US.P6
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1488 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1488
 US-08-867-611-9

Alignment Scores:
 Pred. No.: 1.73 Length: 1488
 Score: 77.50 Matches: 37
 Percent Similarity: 32.60% Conservative: 22
 Best Local Similarity: 20.44% Mismatches: 67
 Query Match: 10.04% Indels: 55
 DB: 3 Gaps: 6

US-09-980-054A-12 (1-148) x US-08-867-611-9 (1-1488)

QY 12 LeuGlulGluGluGlnTyrGlnLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
 Db 436 CTCGACGCTGAAGGATGACATGCTCTCTCCGCCACCATTCTCTGGGATCGTAT 495
 QY 32 GlnTyrLeuGluGlyLysGlu-----Tyr 39
 Db 496 CGTTTTCAGAAAGCGCTTGAACCGTTGGCGATACTCTCGCGTCATCTTGGTATTAT 555
 QY 40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle-----50
 Db 556 GGCTACCGTGGAGGCTTTATCGTGTACGTCACACTGGCAGCCAAAGTCCGTTAGAACAC 615
 QY 51 IleGlyTyrIleGlnLeuMetLysPheTrpLeuSer-----HisIle-----65
 Db 616 ATCGAAATGTTAGACGAGCTTCTGTCTGTGTACGGCGAATAATCCATGTTGCTGTT 675
 QY 66 -----AspGlnTyrAsnAsnSerSer 72
 Db 676 GCTCAGGAAGTTCCTGGCACAGGTGTGGATACCCCTGAAGATCTCGACCCGTCGACGAAT 735
 QY 73 SerLeuArgAsnHisLeuAsn-----79
 Db 736 TCTCTAGACTCCCACTACCGAGGAGTCTGAAGAAGTTAAAGTCGTGCTTCTAAAGTT 795
 QY 80 -----AsnLeuAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluVal 97
 Db 796 AAGCTAACCTGCTGCTGTGAAGAAGCATGCTCTGTGACCCGCCCATCTGCTGCTAAA 855
 QY 98 GluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
 Db 856 TCTAAATTCGGTACCGTCTAAGACAGCTTCGTTGCCAGCTCGTAAAGCTGTTACCCAC 915
 QY 118 ThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThr 137
 Db 916 ATCAACTCTGTTTGGAAA---GATCTGTGGAAGAACAGTTCATCCCGATCGACACCACC 972
 QY 138 Leu 138
 Db 973 ATC 975
 RESULT 14
 PCT-US92-06965A-14

Sequence 14, Application PC/TUS9206965A
 GENERAL INFORMATION:
 APPLICANT: DEVARE, S.
 APPLICANT: DESAI, S.
 APPLICANT: DAILEY, S.
 TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
 NUMBER OF SEQUENCES: 35
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: ABBOTT LABORATORIES
 STREET: ONE ABBOTT PARK ROAD
 CITY: ABBOTT PARK
 STATE: ILLINOIS
 COUNTRY: U.S.
 ZIP: 60065-3500
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/06965A
 FILING DATE: 19920821
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: FOREMSKI, PRISCILLA E.
 REGISTRATION NUMBER: 33,207
 REFERENCE/DOCKET NUMBER: 4834PC.02
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 708-937-6365
 TELEFAX: 708-937-9556
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1488 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: single
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1488
 PCT-US92-06965A-14
 Alignment Scores:
 Pred. No.: 1.73 Length: 1488
 Score: 77.50 Matches: 37
 Percent Similarity: 32.60% Conservative: 22
 Best Local Similarity: 20.44% Mismatches: 67
 Query Match: 10.04% Indels: 55
 DB: 5 Gaps: 6
 US-09-980-054A-12 (1-148) x PCT-US92-06965A-14 (1-1488)
 QY 12 LeuGlulGluGluGlnTyrGlnLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
 Db 436 CTCGACGCTGAAGGATGACATGCTCTCTCCGCCACCATTCTCTGGGATCGTAT 495
 QY 32 GlnTyrLeuGluGlyLysGlu-----Tyr 39
 Db 496 CGTTTTCAGAAAGCGCTTGAACCGTTGGCGATACTCTCGCGTCATCTTGGTATTAT 555
 QY 40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle-----50
 Db 556 GGCTACCGTGGAGGCTTTATCGTGTACGTCACACTGGCAGCCAAAGTCCGTTAGAACAC 615
 QY 51 IleGlyTyrIleGlnLeuMetLysPheTrpLeuSer-----HisIle-----65
 Db 616 ATCGAAATGTTAGACGAGCTTCTGTCTGTGTACGGCGAATAATCCATGTTGCTGTT 675
 QY 66 -----AspGlnTyrAsnAsnSerSer 72
 Db 676 GCTCAGGAAGTTCCTGGCACAGGTGTGGATACCCCTGAAGATCTCGACCCGTCGACGAAT 735
 QY 73 SerLeuArgAsnHisLeuAsn-----79

Db 736 TCTCTAGACTCCCACTACCAGGAGCTTCTGAAAGAGTTAAAGCTGCTCTTCTAAAGTT 795
Qy 80 -----AsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspIysGluVal 97
Db 796 AAGCTAACCTGCTGCTGTGTGAGAGAGCATGCTCTCTGACCCCGCGCACTCTGCTAAA 855
Qy 98 GluAspTyrGluLysAsnIleLysAlaArgAsnLysLeuArgValIleAlaSerIle 117
Db 856 TCTAAATTCGGTTACGGTGTAAAGAGCTTCTGTCACCGCTCGTAAAGCTCTTACCCAC 915
Qy 118 ThrLysGluThrTrpIleAspSerLeuAspAsnLeuValIysGluValGlyGlyThr 137
Db 916 ATCAACTCTGTTTGGAAA---GATCTGCTGAGACACAGCTTACCCCGATCGACACCACC 972
Qy 138 Leu 138
Db 973 ATC 975

RESULT 15
US-08-500-222-1
; Sequence 1, Application US/08500222
; Patent No. 5643733
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESS: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/500,222
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,704
; FILING DATE: 21-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wong, Wean Khing
; REGISTRATION NUMBER: 33561
; REFERENCE/DOCKET NUMBER: 5051.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-9396
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Borrelia burgdorferi
; STRAIN: B31
; IMMEDIATE SOURCE:
; CLONE: PB776
US-08-500-222-1

Alignment Scores:

Pred. No.: 1.75 Length: 1497
Score: 77.50 Matches: 34
Percent Similarity: 35.47% Conservative: 27
Best Local Similarity: 19.77% Mismatches: 46
Query Match: 10.04% Indels: 65
DB: 1 Gaps: 8
US-09-980-054A-12 (1-148) x US-08-500-222-1 (1-1497)
Qy 12 LeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlnIleGlnGlyThrLysAsp 31
Db 436 CTCGACGCTGAAGGATGACACTGTACTTCTCTCGGCCACCATTCCTCGGATCGTGAT 495
Qy 32 GlnTyrLeuGluGlyLysGlu-----Tyr 39
Db 496 CGTTTTCGAGAGGCCCTTGAAACCGTTGGCGATAACTTCTCTGGTCATCTTGGTATTAT 555
Qy 40 GlyTyrGlnThrGlyPheGlnArgPheLeuIle----- 50
Db 556 GGCTACCGTGCAGGCTTTATCCGTCGTTACGTCACTGGCAGCCACAGTCGCTTAGAACAC 615
Qy 51 IleGlyTyrIleGlnGluMetLysPheTrpLeuSer-----HisIle----- 65
Db 616 ATCGAAATGTTAGAGCAGCTTCGTGTTCTGTGTACGGCGAAGAAATCCCATGTGCTGT 675
Qy 65 ----- 65
Db 676 GCTCAGGAAGTTCTCTGGCACAGGTGTGATACCCCTGAAATCCGTCGACAGGCTTATG 735
Qy 66 -----AspGlnTyrAsnAsnSerSerLeu-----ArgAsn 76
Db 736 AAGATCTCAGACCCGAGAAATACTTCAAGGCTATTATTTTATTTCAGACACAGAGGG 795
Qy 77 HisLeuAsnAsnLeuGluAspIle-----MetAlaGlnIleSerIleThrAsnGly 93
Db 796 AATTAAATGAAGTAGAAAAAGTCTTAGTAAGAAATGAAGAAATGGCAGTTCAATCAGGT 855
Qy 94 AspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLysLeuArgVal 113
Db 856 AACGGCACATATTTCAGATGCAGAC-----AGAGGTTCTATACAAAT 897
Qy 114 ---IleAlaSerIleThrLysGluThrTrpIysIle 124
Db 898 GAAATAGAGCACTTACAGACGAAATTAATAGAATT 933

Search completed: February 16, 2004, 16:23:50
Job time : 955 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 15:27:53 ; Search time 332 Seconds
(without alignments)
1642.101 Million cell updates/sec

Title: US-09-980-054A-12
Perfect score: 772
Sequence: 1 MSDDIDNVLNLEBQYELG.....NLVKEVGGLQVSENPDMMW 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlh
-Q=/cn2_1/USPTO_spool/US9980054/runat_13022004_163843_26438/APP_query.fasta_1.327
-DB=publishedApplications_NA -QWIF=ascap -SUFFX=rnpb -MLNMAICH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -RND=0 -THR SCORES=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US9980054 @CN 1.1 391 @runat_13022004_163843_26438
-NCPU=6 -ICPU=3 -NO MWAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOF=6 -FGAPEXT=7 -YGAPOF=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
2: /cn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cn2_6/prodata/1/pubpna/US05_NEW_PUB.seq:
4: /cn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
5: /cn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
6: /cn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
8: /cn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
9: /cn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
13: /cn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:
14: /cn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq:
15: /cn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq:
16: /cn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
17: /cn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
18: /cn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description

1	764	99.0	447	13	US-10-032-585-6062	Sequence 6062, Ap
2	119.5	15.5	2877	15	US-10-128-714-6171	Sequence 6171, Ap
3	119.5	15.5	3863	15	US-10-128-714-171	Sequence 171, App
4	119.5	15.5	4877	15	US-10-128-714-5171	Sequence 5171, Ap
5	107	13.9	65	13	US-10-032-585-2884	Sequence 2884, Ap
6	106.5	13.8	2601	15	US-10-128-714-7171	Sequence 7171, Ap
7	99.5	12.9	430	12	US-10-242-535A-56991	Sequence 56991, A
8	91.5	11.9	3018	12	US-10-369-493-24748	Sequence 24748, A
9	91	11.8	35	13	US-10-032-585-3062	Sequence 3062, Ap
10	87	11.3	5967	15	US-10-171-581-158	Sequence 158, App
11	86	11.1	7100	10	US-09-932-183A-1	Sequence 1, Appl
12	85.5	11.1	6723	10	US-09-070-927A-163	Sequence 163, App
13	85	11.0	255	12	US-10-369-493-44827	Sequence 44827, A
14	83.5	10.8	2637	12	US-10-369-493-44024	Sequence 44024, A
15	83.5	10.8	2640	12	US-10-369-493-24958	Sequence 24958, A
16	83.5	10.8	2643	12	US-10-369-493-45330	Sequence 45330, A
17	83.5	10.8	2816	13	US-10-094-749-289	Sequence 289, App
18	83.5	10.8	3261	15	US-10-005-216-1	Sequence 1, Appl
19	83	10.8	3552	9	US-09-815-242-4174	Sequence 4174, Ap
20	83	10.8	3567	9	US-09-815-242-8027	Sequence 8027, Ap
21	82.5	10.7	1308	13	US-10-032-585-6260	Sequence 6260, Ap
22	82.5	10.7	3678	12	US-10-369-493-45562	Sequence 45562, A
23	82	10.6	507	10	US-09-796-692-4355	Sequence 4355, Ap
24	82	10.6	507	12	US-10-057-475B-4355	Sequence 4355, Ap
25	82	10.6	507	12	US-10-154-884B-4355	Sequence 4355, Ap
26	82	10.6	507	15	US-10-040-862-4355	Sequence 4355, Ap
27	81	10.6	510	11	US-09-918-995-24731	Sequence 24731, A
28	81	10.5	963	9	US-09-815-242-9687	Sequence 9687, Ap
29	81	10.5	3491	8	US-08-781-986A-117	Sequence 117, App
30	80.5	10.4	785	13	US-10-032-585-6511	Sequence 6511, Ap
31	80	10.4	3471	12	US-10-369-493-23730	Sequence 23730, A
32	80	10.4	6773	10	US-09-864-864-336	Sequence 336, App
33	80	10.4	1230025	12	US-10-289-762-1	Sequence 1, Appl
34	79.5	10.3	1319	13	US-09-814-353-21904	Sequence 21904, A
35	79.5	10.3	1934	13	US-09-882-227-201	Sequence 201, App
36	79.5	10.3	68407	12	US-10-398-221-9	Sequence 9, Appl
37	79.5	10.3	3011208	12	US-10-398-221-2058	Sequence 2058, Ap
38	79	10.2	3522	12	US-10-369-493-45641	Sequence 45641, A
39	78.5	10.2	1420	12	US-10-014-099F-62	Sequence 62, Appl
40	78.5	10.2	1441	12	US-10-014-099F-66	Sequence 66, Appl
41	78.5	10.2	2712	12	US-10-369-493-24735	Sequence 24735, A
42	78.5	10.2	5290	12	US-10-014-099F-76	Sequence 76, Appl
43	78.5	10.2	5309	12	US-10-014-099F-77	Sequence 77, Appl
44	78	10.1	1125	13	US-10-032-585-6585	Sequence 6585, Ap
45	77.5	10.0	369	12	US-10-242-535A-42952	Sequence 42952, A

ALIGNMENTS

RESULT 1
US-10-032-585-6062
; Sequence 6062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6662
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6062

Alignment Scores: 9.01e-87 Length: 447
Pred. No.: 764.00 Matches: 146
Score:

Percent Similarity: 100.00%
Best Local Similarity: 98.65%
Query Match: 98.96%
DB: 13
Conservative: 2
Mismatch: 0
Indels: 0
Gaps: 0

US-09-980-054A-12 (1-148) x US-10-032-585-6062 (1-447)

```
QY 1 MetSerAspIleAspIleAspValLeuAsnLeuGluGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATATAGATATAGTATTAAATTTAGAAAGAAACAATATGATAGGA 60
QY 21 PheIysGluGlnIleGlnGlyThrIysAspGlnTyrLeuGluGlyGlyGly 40
DB 61 TTTAAAGAGAGGTCAATACAGGAACAAAGATCAATATTTAGAGAAAGAAATATGAT 120
QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetIysPhe 60
DB 121 TATCAACTGGATTTCAACGATTTTAAATCATTTGTTATATTAACAAGATTAATGAATTT 180
QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerIleuArgAsnHisIleuAsn 80
DB 181 TGGTTATCCCATATAGATCAATATAATACTCTTCTCACTTCGGAATCATTTGAATAAT 240
QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspIysGluValGluAspTyr 100
DB 241 TTGAAATATTTTGGCACAATTTCTATAACGAATGGAGATAAAGAGTTGAAGATTAT 300
QY 101 GluIysAsnIleIysIysAlaArgAsnIysLeuArgValIleAlaSerIleThrIysGlu 120
DB 301 GAAAAAATATTTAAAGGCAAGAAATAAATTAAAGATGATAGCTAGTATAACTAAGAA 360
QY 121 ThrTrpIysIleAspSerLeuAspAsnLeuValIysGluValGlyGlyThrLeuGlnVal 140
DB 361 ACTTGGAAATTCATTCATTAGATATTTGGTGAAGAGTAGTGTGAAGCTTTACAGTT 420
QY 141 SerGluAsnProAspAspMetTrp 148
DB 421 AGTGAAGAACCCGATGATATGTGG 444
```

RESULT 2

US-10-128-714-6171/c

; Sequence 6171, Application US/10128714

; Publication No. US20030119013A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos

; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lemieux, Sebastien M

; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; TITLE OF INVENTION: Methods of Use

; FILE REFERENCE: 10182-018-999

; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6171

; LENGTH: 2877

; TYPE: DNA

; ORGANISM: Aspergillus fumigatus

; US-10-128-714-6171

Alignment Scores:

Pred. No.: 0.00017 Length: 2877

Score: 119.50 Matches: 34

Percent Similarity: 40.67% Conservative: 27

Best Local Similarity: 22.67% Mismatches: 48

Query Match: 15.48% Indels: 41

DB: 3 Gaps: 3

US-09-980-054A-12 (1-148) x US-10-128-714-6171 (1-2877)

```
QY 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheIysGluGln 25
DB 539 CTTGACGCTCTTCGACCTGGAGAGGAATTTACCAAGAGGGCTATAACTTGGGTGCA 480
QY 26 IleGlnGlyThrIysAspGlnTyrLeuGluGlyGlyGlyGlyGlyGlyGly 45
DB 479 ACTGACCGTCCCGGCTGGTTACACTGAAGAGAGCGTGTTCCTGCTGGCCAGAGGCTTC 420
QY 46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetIysPheTrpLeuSerHisIle 65
DB 419 GAAAAGTTTCTTGAATTTGGGAAGACTATATATGGCAAGCACTCGTCTGGCCAGAGGCTC 360
QY 65 ----- 65
DB 359 GCAGACTCTGAATCTCTCGGAATTCGTCAAGAGAGTGAACGGAACAGATACCCGCAA 300
QY 66 ---AspGlnTyrAsn----- 69
DB 299 TCTCAGCATATATAGAGGTGATCTTCTTGGAACCACTTCGTATCAAGAGAGATGCTG 240
QY 70 -----AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
DB 239 CTTCTCCCTCCAGTTCACGCTGGCCAAAACCTCGAGATTTCTACTCGAATTAGTTGAT 180
QY 86 ---AlaGlnIleSerIleThrAsnGlyAspIysGluValGluAspTyrGluIysAsnIle 104
DB 179 CCAGCTTCGTTGCCTATGGAGAACCGGAAGAACTGTACAGATGTTGATGACGCTTT 120
QY 105 LysIysAlaArgAsnIysLeuArgValIle 114
DB 119 AAGGCGCTGCAATCAAGCCCAAGCTTATC 90
```

RESULT 3

US-10-128-714-171/c

; Sequence 171, Application US/10128714

; Publication No. US20030119013A1

; GENERAL INFORMATION:

; APPLICANT: Jiang, Bo

; APPLICANT: Tishkoff, Daniel

; APPLICANT: Zamudio, Carlos

; APPLICANT: Eroshkin, Alexey M

; APPLICANT: Lemieux, Sebastien M

; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

; TITLE OF INVENTION: Methods of Use

; FILE REFERENCE: 10182-018-999

; CURRENT APPLICATION NUMBER: US/10/128,714

; CURRENT FILING DATE: 2002-04-23

; PRIOR APPLICATION NUMBER: US 60/285,697

; PRIOR FILING DATE: 2001-04-23

; PRIOR APPLICATION NUMBER: US 60/287,066

; PRIOR FILING DATE: 2001-04-27

; PRIOR APPLICATION NUMBER: US 60/295,890

; PRIOR FILING DATE: 2001-06-05

; PRIOR APPLICATION NUMBER: US 60/303,899

; PRIOR FILING DATE: 2001-07-09

; PRIOR APPLICATION NUMBER: US 60/316,362

; PRIOR FILING DATE: 2001-08-31

; NUMBER OF SEQ ID NOS: 8603

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 171

; LENGTH: 3863

; TYPE: DNA

```

; ORGANISM: Aspergillus fumigatus
US-10-128-714-171

Alignment Scores:
Pred. No.: 0.000261 Length: 3863
Score: 119.50 Matches: 34
Percent Similarity: 40.67% Conservative: 27
Best Local Similarity: 22.67% Mismatches: 48
Query Match: 15.48% Indels: 41
DB: 15 Gaps: 3

US-09-980-054A-12 (1-148) x US-10-128-714-171 (1-3863)

QY 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
Db 677 CTTGACGGTCTCTCGACCTCGGAAGAGAAATTCACCAAGAGGCGCTATAACTTGGGTGCA 618
QY 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
Db 617 ACTGACGGTCCCGAGGCTGGTTACACTGAAGGAAGCGTGTTCGCGTTGAGAAAGGCTTC 558
QY 46 GlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
Db 557 GAAAAGTTTCTGAATTGGGAAGACTATATGGCAAGCACTCTCTGGGCCAGAGGCTC 498
QY 65 ----- 65
Db 497 GCAGACTCTGAATCTCTCGGAATTCGTCACAGGAGTGCACGGAACAGATACCCGCAAA 438
QY 66 ---AspGlnTyrAsn----- 69
Db 437 TCTCAGCCATATAATGAGGCGTATCTTCTCTTGGAAACCATTCGTATGCAAGGAGATGCTG 378
QY 70 -----AsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
Db 377 CCTCTCCCTCCAGTTCACGGCTGGCCAAAACCTCGAGATTCTACTCGAATTAGTTGAT 318
QY 86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
Db 317 CCAGCTTCGTTGCTATGGAGAACACGGAAGAAGCTGTACAGATGTTGATGAGCGCTT 258
QY 105 LysLysAlaArgAsnLysLeuArgValIle 114
Db 257 AAGGCGCTGCATCAAGGCCAAGCTTATC 228

RESULT 4
US-10-714-5171/c
; Sequence 5171, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wendi
; APPLICANT: Tishkoff, Daniel
; APPLICANT: Zamudio, Carlos
; APPLICANT: Eroshkin, Alexey M
; APPLICANT: Lemieux, Sebastien
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
; TITLE OF INVENTION: Methods of Use
; FILE REFERENCE: 10182-018-999
; CURRENT APPLICATION NUMBER: US/10/128,714
; CURRENT FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 60/285,697
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 5171
; LENGTH: 4877
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
US-10-128-714-5171

Alignment Scores:
Pred. No.: 0.000367 Length: 4877
Score: 119.50 Matches: 34
Percent Similarity: 40.67% Conservative: 27
Best Local Similarity: 22.67% Mismatches: 48
Query Match: 15.48% Indels: 41
DB: 15 Gaps: 3

US-09-980-054A-12 (1-148) x US-10-128-714-5171 (1-4877)

QY 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
Db 1539 CTTGACGGTCTCTCGACCTCGGAAGAGAAATTCACCAAGAGGCGCTATAACTTGGGTGCA 1480
QY 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
Db 1479 ACTGACGGTCCCGAGGCTGGTTACACTGAAGGAAGCGTGTTCGCGTTGAGAAAGGCTTC 1420
QY 46 GlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
Db 1419 GAAAAGTTTCTGAATTGGGAAGACTATATGGCAAGCACTCTCTGGGCCAGAGGCTC 1360
QY 65 ----- 65
Db 1359 GCAGACTCTGAATCTCTCGGAATTCGTCACAGGAGTGCACGGAACAGATACCCGCAAA 1300
QY 66 ---AspGlnTyrAsn----- 69
Db 1299 TCTCAGCCATATAATGAGGCGTATCTTCTCTTGGAAACCATTCGTATGCAAGGAGATGCTG 1240
QY 70 -----AsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
Db 1239 CCTCTCCCTCCAGTTCACGGCTGGCCAAAACCTCGAGATTCTACTCGAATTAGTTGAT 1180
QY 86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
Db 1179 CCAGCTTCGTTGCTATGGAGAACACGGAAGAAGCTGTACAGATGTTGATGAGCGCTT 1120
QY 105 LysLysAlaArgAsnLysLeuArgValIle 114
Db 1119 AAGGCGCTGCATCAAGGCCAAGCTTATC 1090

RESULT 5
US-10-032-585-2884/c
; Sequence 2884, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-008-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2884
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-2884

Alignment Scores:
Pred. No.: 2.59e-05 Length: 65
Score: 107.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0

```


Db 2084 CTGTGATGAAAGAACAGAGCTCTCTATTATCTCTGTTGGAGGAGAGAAACAGAGCAAAA 2143
QY 26 -----IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
Db 2144 CATCTGCCCAACAACTTCACTCTACGACGACAGTTTGTAGTGAGCTTTACCCCTTGGCA 2203
QY 41 TyrGlnThrGlyPheGlnArgPheLeuLeu-----IleGlyTyrIle 54
Db 2204 CAACAGGAGAAAGTTCTCCAAACAACTGATTCAACTACAAAATGACAAACTTAAGATATG 2263
QY 55 GlnGluLeuMetLysPheThrLysSerHisIleAspGlnTyrAsnAsnSerSerLysLeu 74
Db 2264 CARGAAATGAAAATCTCAACAGAACATAGAAATATGGAGACAAACACATTGGAGATG 2323
QY 75 ArgAsnHisLeuAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAsp 94
Db 2324 GAATTAATAATTAAGGGCTTGAAGAGTTAATAAGCACTTTAAAGGATACCAAGAGGCC 2383
QY 95 LysGluValGluAspTyrGluLysAsnIleLysLysAlaArg---AsnLysLeuArgVal 113
Db 2384 CAAAAGGTAACTCACTGGCATATGAATAGAGAACTTCGTCTCAAGAACTTAAACTA 2443
QY 114 IleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGlu 133
Db 2444 AATCGGAATTACTCAAGGATAAAGAGAAATAAATAATTTCAATAACATAATTTCTGAA 2503
QY 134 ValGlyGlyThrLeuGlnValSerGlu 142
Db 2504 TATGAACGTACAACTCAGCAGCTTTGAA 2530

RESULT 11

US-09-932-183A-1
; Sequence 1, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394CI-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 09/308,375
; PRIOR FILING DATE: 1998-05-14
; PRIOR APPLICATION NUMBER: PCT/US98/18828
; PRIOR FILING DATE: 1998-09-08
; PRIOR APPLICATION NUMBER: EP9719636.4
; PRIOR FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7100
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-932-183A-1

Alignment Scores:

Pred. No.:	10, 4	Length:	7100
Score:	86.00	Matches:	35
Percent Similarity:	44.23%	Conservative:	34
Best Local Similarity:	22.44%	Mismatches:	61
Query Match:	11.14%	Indels:	26
DB:	10	Gaps:	6

US-09-980-054A-12 (1-148) x US-09-932-183A-1 (1-7100)

QY 5 AspIleAspAsnValLeuAsnLeu---GluGluGluGlnTyrGluLeuGlyPheLysGlu 23
Db 5677 GATGTCGATAACAAAATTTCAATGACTGAAGAGATGAAGACAGGTTAAATACTATAGC 5736
QY 24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGlyTyrGlnThr 43
Db 5737 AAGCAATAAAGCTCATTCACACAAACAAAGACGAGAGAAATACATTAAGCAGCTT 5796
QY 44 GlyPheGlnArgPheLeuIleIleGlyTyr-----IleGlnGluLeu 57

Alignment Scores:

Db 5797 GAAGACAAAAGAAAGAGCTCGAAAGGTTTCCCTGCATCCAGAACAGATCACTGAGAA 5856
QY 58 MetLysPheThrLysSer-----HisIleAspGlnTyrAsnAsnSerSer 72
Db 5857 ATGCAAACTGGAAAGATAAACAGAAAGATTTTAACTTGAGCTTTATAACACCAAGAAG 5916
QY 73 SerLeuArgAsnHisLeuAsnLeuGluAspIleMetAlaGlnIleSerIleThrAsn 92
Db 5917 TCATCAAGGATATCTATAAATCACTGGCTGATGAAGTTGTATCCATC----- 5964
QY 93 GlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsn---LysLeu 111
Db 5965 -----TACAAAGAGATGTCACGAAAAAATGCTCATATATGAGTTA 6003
QY 112 ArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuVal 131
Db 6004 GAAGCCATCAGAAAGCAGCTCAAGACTTG-----ATCGATGAGATACAGACTGAT 6057
QY 132 LysGluValGlyGlyThrLeuGlnValSerGluAsnProAspAspMet 147
Db 6058 GACGAGGCTAAATTTCAAAAAGAAATTAAGAAAGACAGACAGTATT 6105

RESULT 12

US-09-070-927A-163/c
; Sequence 163, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 163:

US-09-070-927A-163

```
Pred. No.: 11.1 Length: 6723
Score: 85.50 Matches: 46
Percent Similarity: 40.00% Conservative: 20
Best Local Similarity: 27.88% Mismatches: 62
Query Match: 11.08% Indels: 37
DB: 10 Gaps: 6

US-09-980-054A-12 (1-148) x US-09-070-927A-163 (1-6723)

Qy 13 GluGluGluGln-----TyrGluLeuGlyPheLysGluGlyGlnIleGln 27
D5 4382 GAAGAGAGCAAAACGTCGAGAGAAATTCAGTTACAAATAAAGAAATAAGGCCAG 4323
Qy 28 GlyThrLysAspGlnTyLeu-----GluGlyLysGluTyGlyTyGlnThrGly 44
D5 4322 TTGAAGCGGATAGCAAACTCAGCTCTCTGAATTCGAGAAAGAGCATTCAAAATAC 4263
Qy 45 PheGlnArgPheLeuIleGlyTyIleGlnGluLeuMetLysPheTrpLysSerHis 64
D5 4262 AAAGATCGGAAGCCCTAAAAAAGGATCGGCGAGTTAAAGAACAACTATTGATTAA 4203
Qy 65 IleAspGlnTyAsnAsnSerSerLeuArgAsnHisLeu-----AsnAsn 80
D5 4202 ACAGAAGCATATAACATTGCT-----AAAAACACTTGGAAAGTGAAGGAARAC 4152
Qy 81 LeuGluAspIleMetAlaGlnIleSer-----89
D5 4151 TTAGCTGATGAGACGACGAGATTTCTCATTAATAATGACAAATACCATTTGAAGCA 4092
Qy 90 -----IleThrAsnGlyAspLysGluValGluAspTy-----100
D5 4091 ACTGAAAATATTCAACTATCACTAAAGAAATAAAAAATTCGAAGAGTACTACTCAA 4032
Qy 101 -----GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThr 118
D5 4031 AGTAAAGAGCTATTCTAGACAAACTTCGCTAAACAGTGCAGATTACAGAAATAGAT 3972
Qy 119 LysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeu 138
D5 3971 AAAGAAATCGCAATGATTGATGAAAAATAGCCTTGCTAAAGAGTATGAAGACAGCTA 3912
Qy 139 GlnValSerGluAsn 143
D5 3911 TCGATTATTGAAGAT 3897

RESULT 13
; Sequence 44827, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44827
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-44827

Alignment Scores:
Pred. No.: 0.111 Length: 255
Score: 85.00 Matches: 28
Percent Similarity: 48.24% Mismatches: 13

US-09-980-054A-12 (1-148) x US-09-070-927A-163 (1-6723)

Qy 13 GluGluGluGln-----TyrGluLeuGlyPheLysGluGlyGlnIleGln 27
D5 4382 GAAGAGAGCAAAACGTCGAGAGAAATTCAGTTACAAATAAAGAAATAAGGCCAG 4323
Qy 28 GlyThrLysAspGlnTyLeu-----GluGlyLysGluTyGlyTyGlnThrGly 44
D5 4322 TTGAAGCGGATAGCAAACTCAGCTCTCTGAATTCGAGAAAGAGCATTCAAAATAC 4263
Qy 45 PheGlnArgPheLeuIleGlyTyIleGlnGluLeuMetLysPheTrpLysSerHis 64
D5 4262 AAAGATCGGAAGCCCTAAAAAAGGATCGGCGAGTTAAAGAACAACTATTGATTAA 4203
Qy 65 IleAspGlnTyAsnAsnSerSerLeuArgAsnHisLeu-----AsnAsn 80
D5 4202 ACAGAAGCATATAACATTGCT-----AAAAACACTTGGAAAGTGAAGGAARAC 4152
Qy 81 LeuGluAspIleMetAlaGlnIleSer-----89
D5 4151 TTAGCTGATGAGACGACGAGATTTCTCATTAATAATGACAAATACCATTTGAAGCA 4092
Qy 90 -----IleThrAsnGlyAspLysGluValGluAspTy-----100
D5 4091 ACTGAAAATATTCAACTATCACTAAAGAAATAAAAAATTCGAAGAGTACTACTCAA 4032
Qy 101 -----GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThr 118
D5 4031 AGTAAAGAGCTATTCTAGACAAACTTCGCTAAACAGTGCAGATTACAGAAATAGAT 3972
Qy 119 LysGluThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeu 138
D5 3971 AAAGAAATCGCAATGATTGATGAAAAATAGCCTTGCTAAAGAGTATGAAGACAGCTA 3912
Qy 139 GlnValSerGluAsn 143
D5 3911 TCGATTATTGAAGAT 3897

RESULT 13
; Sequence 44827, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44827
; LENGTH: 255
; TYPE: DNA
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-44827

Alignment Scores:
Pred. No.: 0.111 Length: 255
Score: 85.00 Matches: 28
Percent Similarity: 48.24% Mismatches: 13
```

```
Best Local Similarity: 32.94% Mismatches: 34
Query Match: 11.01% Indels: 10
DB: 12 Gaps: 4

US-09-980-054A-12 (1-148) x US-10-369-493-44827 (1-255)

Qy 65 IleAspGlnTyAsnAsnSerSerLeuArgAsnHisLeuAsnLeuGluAspIle 84
D5 1 ATGTCCGAGAGAAAACCTCTGAGATTTCGACTGATTCGATTCGATTCGATTCGATTCG 60
Qy 85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyGlyLysAsnIle 104
D5 61 CTGGAAGAAGTCTCATAGTTCGGTCAAAATCCAAAGGCTGAATTCGAAAAAATACGT 120
Qy 105 LysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTrpLysIle 124
D5 121 ACTAAGCAGAGAAAACCTTGAAA---GATCGCGAGTCACGCTCAGT-----165
Qy 125 AspSerLeuAspAsnLeuVal-----LysGluValGlyGlyThrLeuGln-----139
D5 166 GATGCCAGCGATAAGTTGGTTCGATCAAAAGAAATGGCAGGTCGTGCGGATAATTAT 225
Qy 140 ValSerGluAsnPro 144
D5 226 GTTCTGCAATCCG 240

RESULT 14
US-10-369-493-44024
; Sequence 44024, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44024
; LENGTH: 2637
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-44024

Alignment Scores:
Pred. No.: 5.09 Length: 2637
Score: 83.50 Matches: 37
Percent Similarity: 41.18% Conservative: 40
Best Local Similarity: 19.79% Mismatches: 61
Query Match: 10.82% Indels: 49
DB: 12 Gaps: 6

US-09-980-054A-12 (1-148) x US-10-369-493-44024 (1-2637)

Qy 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGluGlnTyGlnLeuGly 20
D5 1534 TTAAGCAAGATAAATCTCGAAGAT---TTGAAGAGGAGATAAGGAAGATTCAGTTACTA 1590
Qy 21 PheLysGlu-----GlyGlnIleGlnGlyThrLysAspGlnTyTrpLysGlu 35
D5 1591 AAAAGTGAGCAATAAATTAAAGGTTGAAGGTCGAGAGCCCTAAAGGAAGTAATAGG 1650
Qy 36 GlyLysGluTyGlyTyGlnThr-----43
D5 1651 CTTAATGATTATAAAAAATCAATCTACCAAGCTTGAATCGAAATAGATAAGCCAAAAAG 1710
Qy 44 -----GlyPheGlnArgPhe---LeuIle 50
```

```
Db 1711 GAATTATCTGAGATCGAAGATAGGTTACTAAGGTTGGGATTTTAAACAATAGATGAGCTT 1770
Qy 51 IleglyTyriIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
Db 1771 TCTGGGAGATAGGGAATTTGAAAGTTTCACAAATATACATAGAGCAAAAACGCT 1830
Qy 71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle----- 88
Db 1831 GAGAAGGAGTTAAGGACATCCTTGAAGAGTCTTAAAGACGAGAGGAGGAACTTGATAAG 1890
Qy 89 -----SerIleThrAsnGlyAspLys 95
Db 1891 GCCTTCGAAGAACTAGCAAAAATTGAACGGATATAGAAAAAGTCACITTCACAGCTTAAC 1950
Qy 96 GluValGlu-----AspTyrGluLysAsnIleLysAlaArgAsnLysLeuArg 112
Db 1951 GAACCTTCAAAGGAAATTTGACCAAAAAGATACGAAGAAAAAAGGAGAGATGATGAAG 2010
Qy 113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
Db 2011 CTTAGTATGGAGATAAGGATTTGAAACTTAAGTTAGAGAACTTGAGAGAAAGGAT 2070
Qy 133 GluValGlyGlyThrLeuGln 139
Db 2071 GAAATTAATCTACATAGAA 2091
```

RESULT 15

```
US-10-369-493-24958
; Sequence 24958, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24958
; LENGTH: 2640
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-24958
```

```
Alignment Scores:
Pred. No.: 5.1 Length: 2640
Score: 83.50 Matches: 37
Percent Similarity: 41.18% Conservative: 40
Best Local Similarity: 19.79% Mismatches: 61
Query Match: 10.82% Indels: 49
DB: 12 Gaps: 6
```

US-09-980-054A-12 (1-148) x US-10-369-493-24958 (1-2640)

```
Qy 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGly 20
Db 1534 TTAAGCAAGATAAATCTGGAGAT---TTGAAGGGGATAAGGAGATGAGTTACTA 1590
Qy 21 PheLysGlu-----GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGlu 35
Db 1591 AAAAGTGAGACCAATAAATTAAAGGTTGAAGTGGAGCGCTTAAAGGAAAGTAATGAG 1650
Qy 36 GlyLysGluTyrGlyTyrGlnThr----- 43
Db 1651 CTTAATGATTATAAATAATGAATCTACCAAGCTTGAATTCGAATAGATAAGGCCAAAAG 1710
```

```
Qy 44 -----GlyPheGlnArgPhe---LeuIle 50
Db 1711 GAATTATCTGAGATCGAAGATAGGTTACTAAGGTTGGGATTTTAAACAATAGATGAGCTT 1770
Qy 51 IleglyTyriIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
Db 1771 TCTGGGAGATAGGGAATTTGAAAGTTTCACAAATATACATAGAGCAAAAACGCT 1830
Qy 71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle----- 88
Db 1831 GAGAAGGAGTTAAGGACATCCTTGAAGAGTCTTAAAGACGAGAGGAGGAACTTGATAAG 1890
Qy 89 -----SerIleThrAsnGlyAspLys 95
Db 1891 GCCTTCGAAGAACTAGCAAAAATTGAACGGATATAGAAAAAGTCACITTCACAGCTTAAC 1950
Qy 96 GluValGlu-----AspTyrGluLysAsnIleLysAlaArgAsnLysLeuArg 112
Db 1951 GAACCTTCAAAGGAAATTTGACCAAAAAGATACGAAGAAAAAAGGAGAGATGATGAAG 2010
Qy 113 ValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeuValLys 132
Db 2011 CTTAGTATGGAGATAAGGATTTGAAACTTAAGTTAGAGAACTTGAGAGAAAGGAT 2070
Qy 133 GluValGlyGlyThrLeuGln 139
Db 2071 GAAATTAATCTACATAGAA 2091
```

Search completed: February 16, 2004, 17:23:44
Job time : 352 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:25:48 ; Search time 278 Seconds
(without alignments)
1437.111 Million cell updates/sec

Title: US-09-980-054a-12

Perfect score: 772

Sequence: 1 MSDIDIDNVLNLEEQYELG.....NLVKEVGGLQYSENPDDMW 148

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2 1/USPTO.spool/US0980054/runat 13022004.163838 26290/app query.fasta_1.327
-DB=N Geneseq 19Jun03 -CPMT=fstatap -SUFFIX=rng -MINMATCH=0.1 -LOPCL=0
-LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980054 @CGN 1 1 312 @runat 13022004.163838 26290 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:**

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:**
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:**
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:**
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:**
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:**
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:**
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:**
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:**
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:**
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:**
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:**
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:**
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:**
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:**
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:**
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:**
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:**
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:**
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:**
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:**
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:**
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:**
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:**
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:**
- 25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	772	100.0	447	22	AAC86750	DNA encoding a Can
2	764	99.0	447	22	AAS23431	Candida albicans e
3	764	99.0	447	24	ABZ31775	Probe for DNA enco
4	562	72.8	326	22	AAC86756	Aspergillus fumiga
C 5	119.5	15.5	2877	25	ABT20223	Aspergillus fumiga
C 6	119.5	15.5	3863	25	ABT17813	Aspergillus fumiga
C 7	119.5	15.5	4877	25	AST19627	Arabidopsis thalia
8	116	15.0	1372	21	AAC42626	Borrelia burgdorfe
9	107.5	13.9	2462	20	AAZ20299	Candida gene relat
C 10	107	13.9	65	24	ABZ28869	Aspergillus fumiga
C 11	106.5	13.8	2601	25	ABT20821	Human GK003 homolo
12	103.5	13.4	1131	22	ABA09167	Human polynucleoti
13	103.5	13.4	1131	22	AAK52584	Human secreted pro
14	99.5	12.9	445	21	AAC01972	Human colon cancer
15	98.5	12.9	677	24	ABQ56851	Human polynucleoti
16	99.5	12.9	1118	22	AAK51600	DNA encoding novel
17	99.5	12.9	1164	23	AAS92292	Borrelia burgdorfe
C 18	97.5	12.6	111309	20	AAZ20250	Borrelia burgdorfe
C 19	97.5	12.6	910715	20	AAZ20248	Drosophila melanog
20	95.5	12.4	426	23	ABL22081	Drosophila melanog
21	95.5	12.4	2306	23	ABL22078	Drosophila melanog
22	95.5	12.4	2426	23	ABL22080	Drosophila melanog
23	95.5	12.4	3534	23	ABL22082	Drosophila melanog
24	95	12.3	7057	24	ABZ11230	Human polynucleoti
25	94.5	12.2	2991	20	AAZ20288	Borrelia burgdorfe
26	91.5	11.9	1664976	19	AAV21209	Methanococcus jann
C 27	91	11.8	55	24	ABZ28979	Candida gene relat
C 28	91	11.8	80	22	AAS23676	Tetracycline promo
C 29	90.5	11.7	48551	24	AAS20800	Clostridium diffic
30	90	11.7	1428	23	ABL23453	Drosophila melanog
C 31	90	11.7	35515	20	AAZ20252	Borrelia burgdorfe
32	89	11.5	4228	20	AAZ20277	Borrelia burgdorfe
33	88.5	11.5	2422	23	ABLI1444	Drosophila melanog
34	87	11.3	717	20	AAK61834	B. burgdorferi ant
35	87	11.3	828	20	AAK61833	B. burgdorferi ant
36	87	11.3	53585	20	AAZ20251	Borrelia burgdorfe
37	87	11.3	487980	25	AAS32233	Human chromosome 3
C 38	86.5	11.2	1902	20	AAZ20309	Borrelia burgdorfe
39	86	11.1	4731	23	ABL23452	Drosophila melanog
40	86	11.1	7100	20	AAZ24980	Bacillus subtilis
41	85.5	11.1	657	24	ABQ58583	Human colon cancer
C 42	85.5	11.1	6723	20	AAK13100	Enterococcus faeca
C 43	85.5	11.1	6723	24	ABQ58895	Enterococcus faeca
44	85	11.0	840	20	AAZ20354	Borrelia burgdorfe
45	85	11.0	1524	21	AAA70107	Plasmodium falcipa

ALIGNMENTS

RESULT 1

AAC86750

ID AAC86750 standard; DNA; 447 BP.

XX

AC AAC86750;

DT 02-APR-2001 (first entry)

XX

DE DNA encoding a Candida albicans protein CanL260.

XX

KW CaDRA472; CaDR489; 1CaDR527; 2CaDR527; CaFLO24; CanL260; CaDR361;

KW antifungal; fungal infection; pathogenic fungi; ss.

XX

OS Candida albicans.

XX

Key Location/Qualifiers

FT CDS 1..447

```

FT XX /*tag= a
PN WO200075305-A2.
XX
XX 14-DEC-2000.
XX
XX 08-JUN-2000; 2000WO-FR01567.
XX
XX 09-JUN-1999; 99FR-0007250.
XX
XX (HMERI) HOECHST MARION ROUSSEL.
XX
XX Lallane J, Rocher C;
XX
XX WPI; 2001-050024/06.
XX
XX P-PSDB; AAB30859.
XX
XX New polynucleotides from Candida albicans and their derived proteins,
XX useful for diagnosis and treatment of fungal infections and for drug
XX screening.
XX
XX Claim 4; Page 77-78; 89pp; French.
XX
XX The present sequence encodes a Candida albicans protein. The
XX specification describes genes CaDR472, CaDR489, CaDR527,
XX CaFLQ24, CaNL260, and CaDR361. These genes are essential for
XX survival, and so are good targets for antifungal agents. The Candida
XX albicans genes and their derived proteins are used to screen compounds
XX for the ability to inhibit the activity of the protein, i.e. for
XX antifungal activity. The proteins are also used to generate a protective
XX antibody response against fungal infections in mammals. The Candida
XX albicans proteins and genes, and their antibodies, are used for
XX diagnosing fungal infections, specifically C. albicans (in standard
XX amplification, hybridisation or immunological assays, and for studying
XX pathogenic fungi.
XX
XX Sequence 447 BP; 192 A; 38 C; 82 G; 135 T; 0 other.

Alignment Scores:
Pred. No.: 1-79e-80 Length: 447
Score: 772.00 Matches: 148
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-980-054A-12 (1-148) x AAC86750 (1-447)

QY 1 MetSerAspIleAspIleAspValLeuAsnLeuGluGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATAGATATGTTAAATTTAGAGAAGAACAAATATGAAATTAGGA 60
QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGly 40
DB 61 TTAAAGAGGTCAATATACAGGACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120
QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPhe 60
DB 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAGAGATTAATGAAATTT 180
QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
DB 181 TGGTTATCCATATAGATCAATATATAACTCTTCTTCACTTCGGAAATCATTTGATAT 240
QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100
DB 241 TTGGAAGATATTATGGCACAATTTCTATAACGAATGGAGATAAAGAGTTGAAGATTAT 300
QY 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
DB 301 GAAAAAATATTAAAGCGCAGAAATAATTAAGAGTGTAGCTAGTATTAATTAAGAA 360
QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThrLeuGlnVal 140

```

```

DB 361 ACTTGGAAAATGATTCATTTGGATTAATTTGGTGAAGAAGTAGTGGTGAACATTTACAGTT 420
QY 141 SerGluAsnProAspAspMetTyr 148
DB 421 AGTGAAGAACCCCGATGATATGTGG 444

RESULT 2
AAS23431
ID AAS23431 standard; DNA; 447 BP.
XX
XX AAS23431;
XX
XX 04-DEC-2001 (first entry)
XX
XX Candida albicans essential gene CaYNL260C.
XX
XX Gene identification; essential gene; GRACE; pathogenic fungus;
XX gene replacement and conditional expression; fungal infection; ds.
XX
XX Candida albicans.
XX
XX WO200160975-A2.
XX
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05551.
XX
XX 18-FEB-2000; 2000US-0183534.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Roemer T, Jiang B, Boone C, Bussey H;
XX
XX WPI; 2001-489080/53.
XX
XX P-PSDB; AAU5103.
XX
XX Identifying genes essential to fungal metabolism and identifying
XX potential therapeutic agents that target these genes.
XX
XX Claim 22; Page 167; 324pp; English.
XX
XX The present invention relates to novel methods for constructing fungal
XX strains useful for identification and validation of gene products as
XX targets for therapeutic agents, for creating a collection of identified
XX essential genes, and screening assays for the discovery of new drugs.
XX The invention provides the GRACE (gene replacement and conditional
XX expression) method for the construction of mutant organisms referred to
XX as GRACE strains of the organism. The invention can be applied to any
XX organism, particularly a pathogenic fungus e.g. Candida albicans,
XX Aspergillus fumigatus and Cryptococcus neoformans. The methods are
XX useful to identify agents that may be used in the treatment of fungal
XX infections. AAS23381-AAS23442 represent C. albicans essential genes.
XX
XX Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Alignment Scores:
Pred. No.: 1-51e-79 Length: 447
Score: 764.00 Matches: 146
Percent Similarity: 100.00% Conservatives: 2
Best Local Similarity: 98.65% Mismatches: 0
Query Match: 98.96% Indels: 0
DB: 22 Gaps: 0

US-09-980-054A-12 (1-148) x AAS23431 (1-447)

QY 1 MetSerAspIleAspIleAspValLeuAsnLeuGluGluGlnTyrGluLeuGly 20
DB 1 ATGTCAGATATAGATATGATTAATTTAGAGAAGAACAAATATGAAATTAGGA 60
QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGly 40
DB 61 TTAAAGAGGTCAATATACAGGACAAAGATCAATATTTAGAGGAAAGAAATATGTT 120

```

QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnLeuMetLysPhe 60
 DB 121 TATCAAACTGGATTTCAACGATTTTAAATCAATGGTTATATTCAAGAAATTAATGAAATTT 180
 QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerLeuArgAsnHisLeuAsnAsn 80
 DB 181 TGGTTATCCCATAGATCAATATATAACTCTTCTTCACTTCGGAATCAATTTGATATAT 240
 QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspGluValGluAspTyr 100
 DB 241 TTGGAATAATTTTGGCAAAATTTCTATAACGAATGGAGATTAAGAAATTTGAAGATTAT 300
 QY 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
 DB 301 GAAAAAATATTAAGGCAAGAAATAATTAAGGTGATAGCTAGTATATACTTAAGAA 360
 QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
 DB 361 ACTTGAAATTTGATTCATTAGATATTTGGTGAAGAGTAGGTGGAACCTTACAGTT 420
 QY 141 SerGluAsnProAspMetTrp 148
 DB 421 AGTGAACCCCGATGATATGTTGG 444

RESULT 3
 ID ABZ31775
 XX ABZ31775 standard; DNA; 447 BP.
 AC ABZ31775;
 DT 30-JAN-2003 (first entry)
 DE Candida albicans essential gene SEQ ID NO 6062.
 KW Fungus; Yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX Candida albicans.
 XX WO200253728-A2.
 XX 11-JUL-2002.
 PD 26-DEC-2001; 2001WO-US49486.
 XX 29-DEC-2000; 2000US-259128P.
 PR 20-FEB-2001; 2001US-0792024.
 PR 22-AUG-2001; 2001US-314050P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 PI WPI; 2002-566694/60.
 DR P-PSDB; ABP73225.
 XX

Constructing strains for identifying gene products as effective targets for therapeutic intervention, by inactivating in the strain one allele of a gene and placing other allele of the gene under conditional expression -

Claim 37; SEQ ID NO 6062; 167bp + Sequence Listing; English.

The invention relates to constructing (M1) a strain of diploid fungal cells in which both alleles of a gene are modified, comprising modifying one allele by insertion or replacement by a cassette having an expressible selectable marker and modifying other allele by recombination, of a promoter replacement fragment with a heterologous promoter, so that expression of the second allele is regulated by the promoter. (M1) is useful for constructing a strain of diploid fungal cells in which both alleles of a gene are modified. The diploid fungal

CC cells having both alleles modified are useful for identifying a gene that is essential to the survival or growth of a fungus, a gene that contributes to the virulence and/or pathogenicity of a fungus, a gene that contributes to the resistance of a diploid fungus to an antifungal agent, an antifungal agent that inhibits the growth of a diploid fungus and for identifying a therapeutic agent for treatment of a mammalian disease. (M1) is useful for identifying a compound which modulates the activity of a gene product, preferably enzymatic activity, carbon compound catabolism, biosynthetic, transporter, transcriptional, translational, signal transduction, DNA replication and cell division activity. The method is useful for identifying a compound having the ability to inhibit growth or proliferation of C. albicans cells and for treating infection by C. albicans. The present sequence is that of an essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied to Derwent by the European Patent Office.

XX SQ Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Alignment Scores:

Pred. No.: 1,51e-79 Length: 447
 Score: 764.00 Matches: 146
 Percent Similarity: 100.00% Conservatives: 2
 Best Local Similarity: 98.65% Mismatches: 0
 Query Match: 98.96% Indels: 0
 DB: 24 Gaps: 0

US-09-980-054A-12 (1-148) x ABZ31775 (1-447)

QY 1 MetSerAspIleAspIleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGly 20
 DB 1 ATGTCAATATAGATATAGATATATGTTAAATTTAGAGAGACAATATGATTAGGA 60
 QY 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGly 40
 DB 61 TTTAAAGAGCTCAATACAGGACCAAGATCAATATTTAGAGAAAGAAATATGTT 120
 QY 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnLeuMetLysPhe 60
 DB 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGTTATATTCAGAAATTAATGAAATTT 180
 QY 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80
 DB 181 TGGTTATCCCATATAGATCAATATATACTCTTCTTCACTTCGGAATCAATTTGAAATAT 240
 QY 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspGluValGluAspTyr 100
 DB 241 TTGGAATAATTTTGGCAAAATTTCTATAACGAATGGAGATTAAGAAATTTGAAGATTAT 300
 QY 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120
 DB 301 GAAAAAATATTAAGGCAAGAAATAATTAAGGTGATAGCTAGTATATACTTAAGAA 360
 QY 121 ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVal 140
 DB 361 ACTTGAAATTTGATTCATTAGATATTTGGTGAAGAGTAGGTGGAACCTTACAGTT 420
 QY 141 SerGluAsnProAspMetTrp 148
 DB 421 AGTGAACCCCGATGATATGTTGG 444

RESULT 4

AAC86756
 ID AAC86756 standard; DNA; 326 BP.

XX AAC86756;

DT 02-APR-2001 (first entry)

DE Probe for DNA encoding a Candida albicans protein CanL260.

KW CADRA472; CADR489; 1CADR527; 2CADR527; CaFLO24; CanL260; CADR361;

KW antifungal; fungal infection; pathogenic fungi; probe; ss.

XX Candida albicans.

OS WO200075305-A2.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-PR01567.

XX 09-JUN-1999; 99FR-0007250.

XX (HMRI) HOECHST MARION ROUSSEL.

XX Lalanne J, Rocher C;

XX WPI; 2001-050024/06.

XX New polynucleotides from Candida albicans and their derived proteins,
PT useful for diagnosis and treatment of fungal infections and for drug
PT screening

XX Example 5; Page 85-86; 89pp; French.

CC The present sequence represents a probe for DNA encoding a Candida
CC albicans protein. The specification describes genes CADRA472, CADR489,
CC 1CADR527, 2CADR527, CAFLO24, CANL260, and CADR361. These genes are
CC essential for survival, and so are good targets for antifungal agents.
CC The Candida albicans genes and their derived proteins are used to screen
CC compounds for the ability to inhibit the activity of the protein,
CC i.e. for antifungal activity. The proteins are also used to generate a
CC protective antibody response against fungal infections in mammals. The
CC Candida albicans proteins and genes, and their antibodies, are used for
CC diagnosing fungal infections, specifically C. albicans (in standard
CC amplification, hybridisation or immunological assays, and for studying
CC pathogenic fungi.

XX Sequence 326 BP; 146 A; 27 C; 53 G; 100 T; 0 other;

Alignment Scores:

Pred. No.: 2.69e-56 Length: 326
Score: 562.00 Matches: 107
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.07% Mismatches: 0
Query Match: 72.80% Indels: 0
DB: 22 Gaps: 0

US-09-980-054A-12 (1-148) x AAC86756 (1-326)

QY 5 AspIleAspAsnValIleuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGly 24
DB 2 GATATAGATAATGTATTAAATTTAGAGAGATCAATATGATTAGGATTTAAAGAGGT 61
QY 25 GlnIleGlnGlyThrIleAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
DB 62 CAATATACAGGAGAACCAATATATTTAGAGAGAAAGATATGTTATCAACACTGGA 121
QY 45 PheGlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTyrLeuSerHis 64
DB 122 TTTCAACGATTTTATCATCTGGTTATATTCAGATTAATCAATTTGGTTATCCCAT 181
QY 65 IleAspGlnTyrAsnAsnSerSerSerIleuArgAsnHisLeuAsnLeuGluAspIle 84
DB 182 ATAGATCAATATAATAAATCTTTCTTCCTTCGGAATCATTTGAATTAATTTGGAAGATTT 241
QY 85 MetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
DB 242 ATGGGCACAAATTTCTATACGATGGAGATAAGAGATTGAGATTAATGAAAAATATT 301
QY 105 LysLysAlaArgAsnLysLeuArg 112
DB 302 AAAAGGCCAAGAAATAAATAAGA 325

RESULT 5

ID AET20223/c
XX AET20223 standard; DNA; 2877 BP.

XX AC AET20223;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #2581.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.

XX OS Aspergillus fumigatus.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX 23-APR-2002; 2002WO-US13142.

XX 23-APR-2001; 2001US-285697P.

XX 27-APR-2001; 2001US-287066P.

XX 05-JUN-2001; 2001US-295890P.

XX 09-JUL-2001; 2001US-303899P.

XX 31-AUG-2001; 2001US-316362P.

XX (ELIT-) ELITRA PHARM INC.

XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX WPI; 2003-093124/08.

XX New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer

XX Disclosure; Page -: 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case of virulence
CC factors. This polynucleotide sequence represents one of the essential
CC genes of Aspergillus fumigatus of the invention.

XX Sequence 2877 BP; 732 A; 751 C; 744 G; 650 T; 0 other;

Alignment Scores:

Pred. No.: 0.000849 Length: 2877
Score: 119.50 Matches: 34
Percent Similarity: 40.67% Conservative: 27

Best Local Similarity: 22.67% Mismatches: 48
 Query Match: 15.48% Indels: 41
 DB: 25 Gaps: 3

US-09-980-054A-12 (1-148) x ABT20223 (1-2877)

QY 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
 Db 539 CTTGACGGTCTCTCGACTCGAAGAGAAATCTACCAAGAGGGCTATACTTGGGTGCA 480
 QY 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlnThrGlyPhe 45
 Db 479 ACTGACGGTCCCGAGGTGTTACTGAGGAAGCGTGTTCGCGTTGAGAAGGGCTTC 420
 QY 46 GlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
 Db 419 GAAAAGTTCTGAATGGGAAGACTATATGCGCAAGCACTCGTCTGGCCCGCAGAGCTC 360
 QY 65 ----- 65
 Db 359 GCAGACTCGAAATCTCTCGGAATTCGTCACGAGAGTGCAGGAACAGATACCGCAAA 300
 QY 66 ---AspGlnTyrAsn----- 69
 Db 299 TCTCAGCCATATATAGGGGTGATCTCTTGGCAACATTGCGATCGAAGGAGATGCTG 240
 QY 70 -----AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
 Db 239 CCTCTCCCTCCAGTTCACGGCTCGCCCAAAACCTCGAGATTCTACTCGAATTAGTTGAT 180
 QY 86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
 Db 179 CCAGCTTCGTGCTATGAGAGAACCGGAGAGAGCTGTACAGATGTTGATGAGCGCTTT 120
 QY 105 LysLysAlaArgAsnLysLeuArgValIle 114
 Db 119 AAGGCGGTGCAATCAAGCCCAAGCTTATC 90

RESULT 6

ABT17813/C

ID ABT17813 standard; DNA; 3863 BP.

XX AC ABT17813;

XX DT 16-APR-2003 (first entry)

XX DE Aspergillus fumigatus essential gene #171.

XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
 cancer; contamination; biofilm; antibody; immune response; ds.

XX OS Aspergillus fumigatus.

XX PN WO200286090-A2.

XX PD 31-OCT-2002.

XX PF 23-APR-2002; 2002WO-US13142.

XX PR 23-APR-2001; 2001US-285697P.

XX PR 27-APR-2001; 2001US-287066P.

XX PR 05-JUN-2001; 2001US-295890P.

XX PR 09-JUL-2001; 2001US-303899P.

XX PR 31-AUG-2001; 2001US-316362P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;

XX XX WPI; 2003-093124/08.

XX PT New purified or isolated nucleic acids of essential genes of

PT Aspergillus fumigatus, useful for treating or preventing infections by

PT A. fumigatus, or for treating a non-infectious disease in a subject
 PT e.g. cancer
 XX
 PS Disclosure; Page -; 175pp; English.

XX The invention relates to novel purified or isolated nucleic acids of
 CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
 CC the invention are used to treat or prevent infections by a pathogenic
 CC organism such as A. fumigatus, to treat a non-infectious disease in a
 CC subject (e.g. cancer), to prevent or contain contamination of an object
 CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
 CC biofilm comprising A. fumigatus. The polynucleotides are useful for
 CC expressing recombinant protein for characterisation, screening or
 CC therapeutic use, as markers for host tissues in which the pathogenic
 CC organisms invade or reside, for comparing with the DNA sequence of A.
 CC fumigatus to identify duplicated genes or paralogues having the same or
 CC similar biochemical activity and/or function, for comparing with DNA
 CC sequences of other related or distant pathogenic organisms to identify
 CC potential orthologous essential or virulence genes, for selecting and
 CC making oligomers for attachment to a nucleic acid array for examination
 CC of expression patterns, for raising anti-protein antibodies, as an
 CC antigen to raise anti-DNA antibodies or to elicit another immune
 CC response, and for identifying polynucleotides encoding the other protein
 CC with which binding occurs or to identify inhibitors of the binding
 CC interaction. The polypeptides may be used to raise antibodies or to
 CC elicit immune response, as a reagent in assays designed to quantitatively
 CC determine levels of the protein in biological fluids, as a marker for
 CC host tissues in which pathogenic organism invade or reside, and to
 CC isolate correlative receptors or ligands in the case of virulence
 CC factors. This polynucleotide sequence represents one of the essential
 CC genes of Aspergillus fumigatus of the invention.

XX SQ Sequence 3863 BP; 983 A; 980 C; 970 G; 930 T; 0 other;

Alignment Scores:

Pred. No.: 0.00123 Length: 3863
 Score: 119.50 Matches: 34
 Percent Similarity: 40.67% Conservative: 27
 Best Local Similarity: 22.67% Mismatches: 48
 Query Match: 15.48% Indels: 41
 DB: 25 Gaps: 3

US-09-980-054A-12 (1-148) x ABT17813 (1-3863)

QY 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
 Db 677 CTTGACGGTCTCTCGACTCGAAGAGAAATCTACCAAGAGGGCTATACTTGGGTGCA 618
 QY 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlnThrGlyPhe 45
 Db 617 ACTGACGGTGGCCAGGCTGGTACACTGAAGGAAGCGTGTTCGCTGAGAAAGGCTTC 558
 QY 46 GlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIle 65
 Db 557 GAAAAGTTCTGAATGGGAAGACTATATGCGAAAGCACTCGTCTGGGCCAGAGGCTC 498
 QY 65 ----- 65
 Db 497 GCAGACTCGAAATCTCTCGGAATTCGTCACAGGAGAGTGCAGGAACAGATACCGCAAA 438
 QY 66 ---AspGlnTyrAsn----- 69
 Db 437 TCTCAGCCATATATAGGGGTGATCTTCTCTTGGAAACCATTCGTATGCAAGAGATGCTG 378
 QY 70 -----AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMet--- 85
 Db 377 CCTCTCCCTCCAGTTCACGGCTGGCCAAAACCTCGAGATTCCTACTCGAATTAGTTGAT 318
 QY 86 ---AlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIle 104
 Db 317 CCAGCTTCGTGCTATGAGAGACACGAGAGAGCTGTACAGATGTTGATGAGCGCTTT 258
 QY 105 LysLysAlaArgAsnLysLeuArgValIle 114

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141845.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159337.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160778.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 25-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:

Pred. No.: 0.000848 Length: 1372
 Score: 116.00 Matches: 32
 Percent Similarity: 43.90% Conservative: 22
 Best Local Similarity: 26.02% Mismatches: 51
 Query Match: 15.03% Indels: 18
 DB: 21 Gaps: 2

US-09-980-054A-12 (1-148) x AAC42626 (1-1372)

Qy 6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
 Db 185 CTGATTGATTCGTCGTTAGAGGAGACACATGTTCAACAGCGCTTCGATGAGGTTAC 244
 Qy 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGlnTyrGlyGlnThrGlyPhe 45
 Db 245 GAAGAAGCTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 304
 Qy 46 GlnArgPheLeuLeuLeuGlyTyrLeuGlnGluLeuMetLysPheTrpLeuSer 63
 Db 305 GAGACAGCGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 364
 Qy 64 HisLeuAspGlnTyrAsnAsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
 Db 365 CGATTGATCTACACGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 424
 Qy 84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
 Db 425 TTGCTCGATAAAATCCGCTTTTGGATCCGAGGACGAGCTAAAGACGGATCAAGGAT 484
 Qy 104 -----IleLysLysAla 107
 Db 485 GATCTCAGACAGATGTTGAGAGAATGCTTGGTTGTCAGAGGTTTACATATCTCAAGCG 544
 Qy 108 ArgAsnLys 110
 Db 545 CGGAATAAG 553

RESULT 9

AAX20299
 ID AAX20299 standard; DNA; 2462 BP.

XX AC AAX20299;
 XX DT 04-MAY-1999 (first entry)
 XX DE Borrelia burgdorferi polynucleotide sequence #52.
 XX

KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
 KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.

XX WO9858943-A1.

XX 30-DEC-1998.

XX 18-JUN-1998; 98WO-US12764.

XX 03-SEP-1997; 97US-0057483.

XX 20-JUN-1997; 97US-0050359.

XX 22-JUL-1997; 97US-0053344.

XX 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

XX (MEDI-) MEDIMUNE INC.

XX Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO;

XX White OR;

XX WPI; 1999-081217/07.

XX Claim 1; Page 1030-1031; 1128pp; English.

CC AAX20248 to AAX20402 represent polynucleotide sequences isolated from
 CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
 CC the detection, diagnosis, characterisation, prevention and therapy of
 CC Bb infections, e.g. Lyme disease. They can also be used for the
 CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
 CC to a family of motile, spiral-shaped bacteria called Spirochetes.
 CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
 CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
 CC Lyme disease.

SQ Sequence 2462 BP; 1080 A; 229 C; 357 G; 792 T; 4 other;

Alignment Scores:

Pred. No.: 0.0172 Length: 2462
 Score: 107.50 Matches: 46
 Percent Similarity: 43.21% Conservative: 24
 Best Local Similarity: 28.40% Mismatches: 55
 Query Match: 13.92% Indels: 37
 DB: 20 Gaps: 9

US-09-980-054A-12 (1-148) x AAX20299 (1-2462)

Qy 13 GluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGly---ThrLysAsp 31
 Db 1521 AAGCAAGAAGTTATGATTTTATATAAAAAAATGGAAGTTTACTGGGTTTTTATTGAA 1580
 Qy 32 GlnTyrLeuGluGly-----LysGluTyr----- 39
 Db 1581 GAATCTTTGAAAGAGTCAACAGATTTGATTAATAAGCTTTTAAAAAATAAACAATTA 1640
 Qy 40 ---GlyTyrGlnThrGlyPheGlnArg----- 47
 Db 1641 AAAGGATATTAAGAGGATTTTATGAGAAATTTGGTGCACAGAACATATGATATAGAAAGC 1700
 Qy 48 -----PheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSer 63
 Db 1701 ATAAAAATGAATTTTAAACATAGATTTAGTGAAGAGGCAATAGATTTTGTTCCTT 1760
 Qy 64 HisLeuAspGlnTyrAsnAsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
 Db 1761 CATATGATAATTAC---AACTATGAGGTTTAAAGAGAGAAAAATAATTGATGTAGAAAAA 1817

QY 84 IleMetAlaGlnIleSerIleThrAengLyAspLysGluValGluAspTyrGluLysAsn 103
Db 1818 AATTG---CAAAAGACATATCTAGTTAGATCTAAGATAGATAGATAAGTGAAGAAAT 1874
QY 104 -----IleIysLysAlaArgAsnLysLeuArgVal---IleAlaSerIleThr 118
Db 1875 TTAATGTTAAGATAGATAATCTAGAGAGAATTTAAATTTAAATAGATAGTGTAAA 1934
QY 119 LysGlu---ThrTrpLysIleAspSerLeuAspAsnLeuValLysGluValGlyThr 137
Db 1935 AATGAATTAATCTTAAATAGATAGTCTAGATAGATAGATAGATAGTGAAGAAACT 1994
QY 138 LeuGln 139
Db 1995 TTGCAA 2000
RESULT 10
ID ABZ28869/c
XX ABZ28869 standard; DNA; 65 BP.
AC ABZ28869;
XX
XX 30-JAN-2003 (first entry)
DT
XX
XX Candida gene related tetracyclin promoter PCR primer SEQ ID NO 2884.
DE
XX
XX Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
XX
XX Candida albicans.
OS
XX WO200253728-A2.
PN
XX
XX 11-JUL-2002.
PD
XX
XX 26-DEC-2001; 2001WO-US49486.
PF
XX
XX 29-DEC-2000; 2000US-259128P.
PR 20-FEB-2001; 2001US-0792024.
PR 22-AUG-2001; 2001US-314050P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
PI
XX
XX WPI; 2002-566694/60.
DR
XX
XX Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele
PT of a gene and placing other allele of the gene under conditional
PT expression
XX
XX Claim 76; SEQ ID NO 2884; 167pp + Sequence Listing; English.
PS
XX
XX The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division

CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of a PCR
CC primer used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office.
XX
SQ Sequence 65 BP; 20 A; 11 C; 2 G; 32 T; 0 other;
Alignment Scores:
Pred. No.: 0.000198 Length: 65
Score: 107.00 Matches: 21
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.86% Indels: 0
DB: 24 Gaps: 0
US-09-980-054a-12 (1-148) x ABZ28869 (1-65)
QY 2 SerAspIleAspIleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPhe 21
Db 65 TCAGATATAGATATAGATATAGTATTAAATTAGAGAGACATATAGATTAGGATTT 6
QY 22 Lys 22
Db 5 AAA 3
RESULT 11
ABT20821/c
ID ABT20821 standard; DNA; 2601 BP.
XX
XX AC ABT20821;
XX
XX 16-APR-2003 (first entry)
DT
XX
XX Aspergillus fumigatus essential gene #3179.
DE
XX
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response; ds.
KW
XX
XX Aspergillus fumigatus.
OS
XX WO200286090-A2.
PN
XX
XX 31-OCT-2002.
PD
XX
XX 23-APR-2002; 2002WO-US13142.
PF
XX
XX 23-APR-2001; 2001US-285697P.
PR 27-APR-2001; 2001US-287066P.
PR 05-JUN-2001; 2001US-295890P.
PR 09-JUL-2001; 2001US-303899P.
PR 31-AUG-2001; 2001US-316362P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PI
XX
XX WPI; 2003-093124/08.
DR
XX
XX New purified or isolated nucleic acids of essential genes of
PT Aspergillus fumigatus, useful for treating or preventing infections by
PT A. fumigatus, or for treating a non-infectious disease in a subject
PT e.g. cancer
XX
XX Disclosure; Page -; 175pp; English.
PS
XX
XX The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object

CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a cDNA encoding a
CC novel human polypeptide of the invention.

XX SQ Sequence 1131 BP; 340 A; 266 C; 253 G; 270 T; 2 other;

Alignment Scores:
Pred. No.: 0.0187 Length: 1131
Score: 103.50 Matches: 43
Percent Similarity: 44.38% Conservative: 32
Best Local Similarity: 25.44% Mismatches: 60
Query Match: 13.41% Indels: 34
DB: 22 Gaps: 7

US-09-980-054A-12 (1-148) x ABA09167 (1-1131)

QY 11 AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30
Db AACATGCAAGACGAGTCAAGAA---GGTTATAGATGGAATAGATGCTGGCAAGCA 417
QY 31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuLeu 50
Db GTTACTCTTCAACAGGCTTCAATCAAGGTATAGAAAGGTGCAAGAGTCATTTTAAAC 477
QY 51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70
Db TATGGACGACTCCGAGGACATTCAGTGCTTCTCTCTGGTGTCACCTTCATATATAT 537
QY 71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle----- 88
Db AATTCACTTTGATCAATAAAATAAACAATCTTCTGGATGTCAGTTGCCAGGTGTGAAGAG 597
QY 89 -----SerIleThrAsnGlyAsp-----Lys 95
Db TATGTGCTCAACATCTGAATCAATCACTCCACCGTCCCATCTTGTAGATTATTGGAC 657
QY 96 GluValGluAspTyr-----GluLysAsnIleLysLysAla 107
Db TCCATTGAGGATATGGACCTTTGTGATGTATGTTCCAGCTGAGAAAGATTTGATGAAGCT 717
QY 108 ArgAsn-Lys-----LeuArgValIleAlaSerIleThrLysGluThrTrpLys----- 123
Db AAGATGAAAGACTCTGTGAATAATATGCTGAGTTTAAACAAACTGTAGCAGAGCCAT 777
QY 124 -----IleAspSerLeuAspAsnLeuValLysGluValGlyGlyThrLeuGlnVa 140
Db AGTGGGATAGATTGTTTCATATGTAGATGTGTGTAGAACAC-----AGGAGCATGCACAT 831
QY 140 lSerGluAsnProAspAspMetTrp 148
Db TCAGGAAACCCAGCCCATGG 856

RESULT 13

AAK52584

ID AAK52584 standard; cDNA; 1131 BP.

XX AC AAK52584;

XX DT 06-NOV-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 2113.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation; ss.

OS Homo sapiens.

XX PN WO200157190-A2.

XX

PD 09-AUG-2001.

XX PF 05-FEB-2001; 2001WO-US04098.

XX PR 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

PR 20-JUN-2000; 2000US-0598075.

PR 19-JUL-2000; 2000US-0620325.

PR 01-SEP-2000; 2000US-0654936.

PR 15-SEP-2000; 2000US-0663561.

PR 20-OCT-2000; 2000US-0693325.

XX PR 30-NOV-2000; 2000US-0728422.

PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;

PI Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;

PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

XX WPI; 2001-476283/51.

DR P-PSDB; AAM79451.

XX

PT Nucleic acids encoding polypeptides with cytokine-like activities,

PT useful in diagnosis and gene therapy -

XX

PS Claim 1; Page 4491; 6221pp; English.

XX

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the

CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666

CC (AAM80020) are omitted as the relevant pages from the sequence listing

CC were missing at the time of publication.

XX

SQ Sequence 1131 BP; 340 A; 266 C; 253 G; 270 T; 2 other;

Alignment Scores:

Pred. No.: 0.0187 Length: 1131

Score: 103.50 Matches: 43

Percent Similarity: 44.38% Conservative: 32

Best Local Similarity: 25.44% Mismatches: 60

Query Match: 13.41% Indels: 34

DB: 22 Gaps: 7

US-09-980-054A-12 (1-148) x AAK52584 (1-1131)

QY 11 AsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys 30

Db AACATGCAAGACGAGTCAAGAA---GGTTATAGATGGAATAGATGCTGGCAAGCA 417

QY 31 AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuLeu 50

Db GTTACTCTTCAACAGGCTTCAATCAAGGTATAGAAAGGTGCAAGAGTCATTTTAAAC 477

QY 51 IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn 70

Db TATGGACGACTCCGAGGACATTCAGTGCTTCTCTCTGGTGTCACCTTCATATATAT 537

QY 71 SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIle----- 88

Db AATTCACTTTGATCAATAAAATAAACAATCTTCTGGATGTCAGTTGCCAGGTGTGAAGAG 597

QY 89 -----SerIleThrAsnGlyAsp-----Lys 95

Db TATGTGCTCAACATCTGAATCAATCACTCCACCGTCCCATCTTGTAGATTATTGGAC 657

CC nucleic acid which hybridises to (I) in a cell. A probe/primer derived
CC from (I) can be used for determining the presence of a nucleic acid which
CC hybridises to (I), and for determining the phenotype of cells in a sample
CC of cells from a patient. (I) is useful for determining the presence of
CC colon cancer in a cell or tissue type, for determining the presence or
CC state of other type of cancer, in antisense therapy, to generate
CC macroarrays on a solid surface, to identify a chromosome on which the
CC corresponding gene resides, and in tissue profiling, forensics, genetic
CC analysis, mapping and diagnostic applications. (I) can be used to raise
CC antibodies, and to screen for peptide analogues and antagonists.

XX
SQ Sequence 677 BP; 215 A; 129 C; 166 G; 165 T; 2 other;

Alignment Scores:

Pred. No.:	0.0284	Length:	677
Score:	99.50	Matches:	32
Percent Similarity:	45.83%	Conservative:	23
Best Local Similarity:	26.67%	Mismatches:	46
Query Match:	12.89%	Indels:	19
DB:	24	Gaps:	3

US-09-980-054A-12 (1-148) x ABQ56851 (1-677)

Qy	11	AsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLys	30
Db	129	AACATGCAAGACGAGTCAAAGAA--GGTTATAGATGGAATAGATGCTGCAAGCA	185
Qy	31	AspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle	50
Db	186	GTTACTCTTCAACAGGCTTCATCAAGTTATAGAAAGTGCAGAGTCATTTTAAC	245
Qy	51	IleGlyTyrIleGlnGluLeuMetLysPheTrpLeuSerHisIleAspGlnTyrAsnAsn	70
Db	246	TATGGACGACTCCGAGGACATTCAGTCTTCTCTCTGCTGTCACCTTCATAATAAT	305
Qy	71	SerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSerIle	90
Db	306	AATTCACCTTTGATCAATAAATAAACAATCTTCGTGATGCTGCTGCCAG-----	356
Qy	91	ThrAsnGlyAspLysGluValGluAspTyrGluLysAsnIleLysLysAlaArgAsnLys	110
Db	357	-----TGTGAAGAGTAT-----GTG	371
Qy	111	LeuArgValIleAlaSerIleThrLysGluThrTrpLysIleAspSerLeuAspAsnLeu	130
Db	372	CTCAAAACATCGAATCAATCACTCCACCGTCCCTGTTGTAGATTATTGGACTCCATT	431

Search completed: February 16, 2004, 14:33:36
Job time : 289 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:24:58 ; Search time 33 Seconds
(without alignments)
939.048 Million cell updates/sec

Title: US-09-980-054A-12

Perfect score: 772

Sequence: 1 MSDDIDNVLNLEEEQYELG.....NLVKEVGTTQVSENPDMDW 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/FCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	764	99.0	148	12	US-10-032-585-7062
2	91.5	11.9	1005	12	US-10-369-493-1061
3	86.5	11.2	876	15	US-10-198-070-56
4	86.5	11.2	930	15	US-10-198-070-65
5	86	11.1	2285	10	US-09-932-183A-2
6	85	11.0	85	12	US-10-369-493-21140
7	83.5	10.8	688	12	US-10-094-749-1928
8	83.5	10.8	876	15	US-10-198-070-78
9	83.5	10.8	879	12	US-10-369-493-1271
10	83.5	10.8	879	12	US-10-369-493-20337
11	83.5	10.8	880	12	US-10-369-493-21643
12	83.5	10.8	930	12	US-10-153-244-102
13	83.5	10.8	930	12	US-10-210-152-19
14	83.5	10.8	930	15	US-10-005-216-2
15	83.5	10.8	930	15	US-10-198-070-73

16	83.5	10.8	930	15	US-10-198-070-102
17	83	10.8	1184	9	US-09-815-242-5229
18	83	10.8	1188	9	US-09-815-242-12125
19	82.5	10.7	435	12	US-10-032-585-7260
20	82.5	10.7	1225	12	US-10-369-493-21875
21	82	10.6	1164	12	US-10-369-493-6564
22	82	10.6	1963	12	US-10-369-493-5307
23	82	10.6	1963	12	US-10-369-493-5308
24	81	10.5	320	9	US-09-815-242-13785
25	80.5	10.4	254	12	US-10-032-585-7511
26	80	10.4	235	12	US-10-383-834-5
27	80	10.4	457	12	US-10-369-493-5960
28	80	10.4	1156	12	US-10-369-493-43
29	80	10.4	1156	12	US-10-289-762-171
30	79.5	10.3	600	12	US-09-892-227-202
31	79	10.2	1173	12	US-10-369-493-21954
32	78.5	10.2	472	12	US-10-014-099F-63
33	78.5	10.2	479	12	US-10-014-099F-67
34	78.5	10.2	903	12	US-10-369-493-1048
35	78	10.1	374	12	US-10-032-585-7585
36	77.5	10.0	1143	12	US-10-369-493-11081
37	77.5	10.0	1173	12	US-10-369-493-5025
38	77	10.0	267	9	US-09-815-242-11549
39	77	10.0	320	9	US-09-815-242-10422
40	77	10.0	320	12	US-10-369-493-23638
41	77	10.0	517	9	US-09-815-242-5387
42	77	10.0	519	9	US-09-815-242-12331
43	77	10.0	519	9	US-09-815-242-12869
44	77	10.0	529	12	US-10-301-997-74
45	77	10.0	1572	15	US-10-037-182-20

ALIGNMENTS

RESULT 1

US-10-032-585-7062
; Sequence 7062, Application US/10032585
; Publication NO. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7062
; LENGTH: 148
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7062

Query Match	99.0%	Score 764	DB 12	Length 148
Best Local Similarity	98.6%	Pred. No. 1.5e-69		
Matches 146	Conservative 2	Mismatches 0	Indels 0	Gaps 0
QY	1	MSDDIDNVLNLEEEQYELGFKGQIQGTQKQYLEGKEYGYQTGFQRFLLIYGIQELMKF	60	
Db	1	MSDDIDNVLNLEEEQYELGFKGQIQGTQKQYLEGKEYGYQTGFQRFLLIYGIQELMKF	60	
QY	61	WLSHIDQYNNSSSLNHLNLEIDMAQISITNGDKVEDYKNIKKARKLVIASITKE	120	
Db	61	WLSHIDQYNNSSSLNHLNLEIDMAQISITNGDKVEDYKNIKKARKLVIASITKE	120	
QY	121	TWKIDSLNVLKVEGGTTLQVSENPDMDW	148	
Db	121	TWKIDSLNVLKVEGGTTLQVSENPDMDW	148	

Db	723	LFKQWSELQGHKGFQDAEMKNENBEKKGILGSHEDLSKFSLDRNQLAHNKQSST	78.1
Qy	71	SSSLRNHLNLED-----IMAQISITNGDKEVEDYE-----KNIKKARN	109
Db	783	RSSEDFHLNSFNSPPROYQKIMKRLIKRYVLQAI-----DKESDEVNEGELKEIKQDIS	837
Qy	110	KLRVIASITKETWKIDSLDNLVKEVGGTLOVSENPD	146
Db	838	SLR-YELLEKSQNTDLAELIRKLGERLSLESQEE	873
RESULT 4			
US-10-198-070-65			
; Sequence 65, Application US/10198070			
; Publication No. US20030109437A1			
; GENERAL INFORMATION:			
; APPLICANT: AVERBACK, PAUL			
; APPLICANT: GEMMELL, JACK			
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER			
; TITLE OF INVENTION: CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF			
; TITLE OF INVENTION: CELLS			
; FILE REFERENCE: 59003.000008			
; CURRENT APPLICATION NUMBER: US/10/198,070			
; CURRENT FILING DATE: 2002-07-19			
; PRIOR APPLICATION NUMBER: 60/306,161			
; PRIOR FILING DATE: 2001-07-19			
; PRIOR APPLICATION NUMBER: 60/306,150			
; PRIOR FILING DATE: 2001-07-19			
; PRIOR APPLICATION NUMBER: 60/331,477			
; PRIOR FILING DATE: 2001-11-16			
; NUMBER OF SEQ ID NOS: 125			
; SOFTWARE: Patent In Ver. 2.1			
; SEQ ID NO 65			
; LENGTH: 930			
; TYPE: PRT			
; ORGANISM: Rattus norvegicus			
US-10-198-070-65			
Query Match 11.2%; Score 86.5; DB 15; Length 930;			
Best Local Similarity 24.2%; Pred. No. 6;			
Matches 38; Conservative 29; Mismatches 55; Indels 35; Gaps 6			
Qy	19	LGFGKQ---TGCTKDQVLEKGEYGTGTFQRFLLIGVIGELMKF-----WLSHIQYNN	70
Db	777	LFKQWSELQGHKGFQDAEMKNENBEKKGILGSHEDLSKFSLDRNQLAHNKQSST	836
Qy	71	SSSLRNHLNLED-----IMAQISITNGDKEVEDYE-----KNIKKARN	109
Db	837	RSSEDFHLNSFNSPPROYQKIMKRLIKRYVLQAI-----DKESDEVNEGELKEIKQDIS	891
Qy	110	KLRVIASITKETWKIDSLDNLVKEVGGTLOVSENPD	146
Db	892	SLR-YELLEKSQNTDLAELIRKLGERLSLESQEE	927
RESULT 5			
US-09-932-183A-2			
; Sequence 2, Application US/09932183A			
; Patent No. US20020127641A1			
; GENERAL INFORMATION:			
; APPLICANT: Estell, David A.			
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms			
; FILE REFERENCE: GC394CI-US			
; CURRENT APPLICATION NUMBER: US/09/932,183A			
; CURRENT FILING DATE: 2001-08-17			
; PRIOR APPLICATION NUMBER: US 09/308,375			
; PRIOR FILING DATE: 1999-05-14			
; PRIOR APPLICATION NUMBER: PCT/US98/18828			
; PRIOR FILING DATE: 1998-09-08			
; PRIOR APPLICATION NUMBER: EP9719636.4			
; PRIOR FILING DATE: 1997-09-15			
; NUMBER OF SEQ ID NOS: 3			

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2285
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-932-183A-2

Query Match      11.1%; Score 86; DB 10; Length 2285;
Best Local Similarity 22.4%; Pred. No. 22;
Matches 35; Conservative 34; Mismatches 61; Indels 26; Gaps 6;

QY 5 DIDVNLN-BEEQYELGFKGQIGTKDQYLEGKEYGYGTGQRFLLIIGY-----IQEL 57
Db 1856 DVDNKISMTDEDEKVKYISKQILIQOQKQKAKYIKOLEQKKAAGFPDIQEQITEE 1915
QY 58 MKFWLS-----HIDQYNSSSLRNHLNLEDIMAQISITNGDKVEDEYENKIKARN-KL 111
Db 1916 MQWKDKQKQPNLNYTKSIKDIYKSLADEVVISI-----YKMYEKMRDIEL 1964
QY 112 RVASITKWTWKIDSLNVLNKEVGTTQVSENPDPM 147
Db 1965 EAHOKATQDL--IDEIDKTDDEAKFQELKERQDSI 1998

RESULT 6
US-10-369-493-21140
; Sequence 21140, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21140
; LENGTH: 85
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-369-493-21140

Query Match      11.0%; Score 85; DB 12; Length 85;
Best Local Similarity 32.9%; Pred. No. 0.37;
Matches 28; Conservative 13; Mismatches 34; Indels 10; Gaps 4;

QY 65 IDQYNSSSLRNHLNLEDIMAQISITNGDKVEDEYENKIKARNKLRVIASITKETWKI 124
Db 1 MSQKNSEDURTELQSLADTLEEVSNGSKAELEKRSKAEKTLK-DARVTLG----- 55

QY 125 DSLNLY---KEVGTTQ--VSENP 144
Db 56 DASDKLVDTQKEMAGRADNVDRNP 80

RESULT 7
US-10-094-749-1928
; Sequence 1928, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; FILE REFERENCE: 59003.000008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 78
; LENGTH: 876
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-198-070-78

Query Match      10.8%; Score 83.5; DB 15; Length 876;
Best Local Similarity 25.3%; Pred. No. 11;
```

```

; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYUKI
; APPLICANT: NAGAHARA, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1928
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-1928

Query Match      10.8%; Score 83.5; DB 12; Length 688;
Best Local Similarity 22.2%; Pred. No. 8.1;
Matches 35; Conservative 32; Mismatches 48; Indels 43; Gaps 7;

QY 5 DIDVNLNLEEQYELGFKGQIGTKDQYLEGKEYGYGTGQRFLLIIGYIQELMKFWLSH 64
Db 126 DDETVRNL-QEQLQANQE-----KTOAVE-----LWQTVSQE-----LDRHLKLYOE 168
QY 65 IDQYV---NSSSLRNHLNLEDIMAQISITNGDKVE-----EDYEKNI 104
Db 169 MTEAQIHVFESQKQKQDQDFQQLTKQLHVTNENNEVNTNQQLTKVTTCQSVIIEQLRK 228
QY 105 KARNKLRVIASITKETWKIDSLNVLNKEVGTTQVSE 142
Db 229 ROAKLELRVAVA-----KVEELTNVTEDLQGMKKKE 260

RESULT 8
US-10-198-070-78
; Sequence 78, Application US/10198070
; Publication No. US20030109437A1
; GENERAL INFORMATION:
; APPLICANT: AVERBACK, PAUL
; APPLICANT: GEMMELL, JACK
; TITLE OF INVENTION: PEPTIDES EFFECTIVE IN THE TREATMENT OF TUMORS AND OTHER
; CONDITIONS REQUIRING THE REMOVAL OR DESTRUCTION OF
; FILE REFERENCE: 59003.000008
; CURRENT APPLICATION NUMBER: US/10/198,070
; CURRENT FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: 60/306,161
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/306,150
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/331,477
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 78
; LENGTH: 876
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-198-070-78

Query Match      10.8%; Score 83.5; DB 15; Length 876;
Best Local Similarity 25.3%; Pred. No. 11;
```


! ORGANISM: Mus musculus
US-10-198-070-73

Query Match 10.8%; Score 83.5; DB 15; Length 930;
Best Local Similarity 25.3%; Pred. No. 12;
Matches 37; Conservative 29; Mismatches 49; Indels 31; Gaps 6;
QY 19 LGFKEGO---IQGTDQYLEGKEYGYGTGFORFLIIGYIQELMKF-----WLSHIDQYNN 70
Db 777 LKFKKWCCELIQKQKQFQDAEMNKRNEKFKGISGSHEDLSKFSLDKKNQLAHNKQSST 836
QY 71 SSSLRNHLNLEDIMAQISITNGDKEVEDYERNIKKRNKLRYIASITKWTKIDSLDNL 130
Db 837 RSSDYHLN-----SFSNPPRQ---YQIMKRLIKRYVLQAIQDKESDEVN--EGE 882
QY 131 VKEVGCTL-----QVSENPDDM 147
Db 883 LKEIKODISSRLRYELLEKSQNTEDL 908

Search completed: February 16, 2004, 14:28:33
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2004, 14:12:37 ; Search time 28 Seconds
(without alignments)
223.643 Million cell updates/sec

Title: US-09-980-054A-12

Perfect score: 772

Sequence: 1 MSDDIDNVLNLEEEQYELG.....NLVKEVGTTQVSNPDDMW 148

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listings first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/iaa/6C_COMB.pep.*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	86	11.1	2285	4	US-09-308-375-2
2	82	10.6	756	4	US-09-134-001C-3553
3	81	10.5	521	4	US-09-508-370A-3
4	80.5	10.4	275	2	US-08-392-625-17
5	80.5	10.4	275	2	US-08-466-961A-17
6	80	10.4	1156	4	US-09-198-452A-171
7	78	10.1	506	4	US-09-252-991A-23066
8	78	10.1	1104	3	US-08-923-992A-4
9	77.5	10.0	496	3	US-08-867-611-10
10	77.5	10.0	496	5	PCR-US92-06965A-15
11	77.5	10.0	498	1	US-08-500-222-2
12	77.5	10.0	498	1	US-08-500-125-2
13	77.5	10.0	498	2	US-07-779-704B-2
14	77	10.0	425	3	US-08-867-611-22
15	77	10.0	425	5	PCR-US92-06965A-27
16	77	10.0	497	4	US-09-328-352-7286
17	77	10.0	529	4	US-08-887-534A-74
18	77	10.0	529	4	US-09-527-431-74
19	77	10.0	617	3	US-08-867-611-30
20	77	10.0	617	5	PCR-US92-06965A-35
21	76.5	9.9	496	3	US-08-926-842B-12
22	76.5	9.9	810	4	US-09-540-824-25
23	76	9.8	396	1	US-08-430-024-2
24	76	9.8	396	1	US-08-782-003-2
25	76	9.8	396	3	US-09-017-302-2
26	76	9.8	1164	3	US-08-923-992A-2
27	75.5	9.8	258	4	US-09-107-532A-4489
28	75.5	9.8	330	4	US-09-134-001C-3811
29	75.5	9.8	1211	4	US-09-134-001C-4820
30	75	9.7	1010	4	US-09-134-001C-5178
31	74.5	9.7	255	4	US-08-858-207A-465
32	74.5	9.7	263	1	US-08-160-670A-32
33	74.5	9.7	264	1	US-08-160-670A-28
34	74.5	9.7	488	1	US-08-190-802A-60
35	74.5	9.7	488	3	US-08-477-346-60
36	74.5	9.7	488	4	US-08-473-089-60
37	74.5	9.7	488	4	US-08-487-072A-60
38	74.5	9.7	527	1	US-08-160-670A-8
39	74.5	9.7	528	1	US-08-160-670A-5
40	74	9.6	466	2	US-08-912-129A-55
41	73.5	9.5	244	2	US-08-553-633A-5
42	73.5	9.5	580	4	US-09-256-000-21
43	73.5	9.5	593	2	US-08-987-466-1
44	73.5	9.5	593	3	US-09-240-359-1
45	73.5	9.5	983	4	US-09-134-001C-3814

ALIGNMENTS

RESULT 1

US-09-308-375-2

; Sequence 2, Application US/09308375

; Patent No. 6300117

; GENERAL INFORMATION:

; APPLICANT: Genencor International, Inc.

; TITLE OF INVENTION: Proteases From Gram-Positive Organisms

; FILE REFERENCE: GC394-PC7

; CURRENT APPLICATION NUMBER: US/09/308,375

; CURRENT FILING DATE: 1999-05-14

; EARLIER APPLICATION NUMBER: EP9719636.4

; EARLIER FILING DATE: 1997-09-15

; NUMBER OF SEQ ID NOS: 3

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 2285

; TYPE: PRT

; ORGANISM: Bacillus subtilis

US-09-308-375-2

Query Match

Best Local Similarity 11.1%; Score 86; DB 4; Length 2285;

Matches 35; Conservative 34; Mismatches 61; Indels 26; Gaps 6;

QY 5 DIDNVLNL-BEEQVELGFGKEGQIGTKDQYLEGKEYGTGTFQRFLLIGY-----IQEL 57

Db 1856 DVDNKNISMTEDDKVKYYSKQIKLIQQQKQKAKYIKQLEQKKAAGFPDIQEITEE 1915

QY 58 MKFWLS-----HIDQNNSSSLNHLNLEDMQAQISITNGKVEDYENKIKARN-KL 111

Db 1916 MNQWQKQKQDFNLNLTNKKSIKDIYKSLADEVVS-----YKEMYSKMRDIEL 1964

QY 112 RVIASITKTKWTKIDSLNVLNKEVGTTQVSNPDDM 147

Db 1965 EAHQKAQDL--IDEIDKTDDEAKFKQELKQKQDSI 1958

US-09-134-001C-3553

; Sequence 3553, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lyvin Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

US-09-134-001C-3553

; Sequence 3553, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lyvin Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

```
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3553
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3553

Query Match          10.6%; Score 82; DB 4; Length 756;
Best Local Similarity 23.0%; Pred. No. 2.3;
Matches 35; Conservative 31; Mismatches 58; Indels 28; Gaps 6;

QY 7 DNVLNLEEEQVELGFKGQIGTKDQYLEGKEYGVQ-----TGQRFLLIIGYIOELMKFW 61
Db 217 EDVIRLELSE-----QYSLQDLKELGQYGFDIRPATNFK-----AVQWLYAY 265

QY 62 LSHIDQYNNSSSLRNHLNLEDDIMAQISITNGKDEVEDYEKNIKAKRNKLRLVIA-SITKE 120
Db 266 LAAIKEQGAAMSLGRTSTFLDIYAERDLQNGDITEQEVQEIIDHFIMKLRIKRVKFAKTP 325

QY 121 -----TWKIDSLDLNVLKVEGVTLOVSEN 143
Db 326 YNELFSGDPTWTESIGGV--GIDRPMVTKN 355

RESULT 3
US-09-508-370A-3
; Sequence 3, Application US/09508370A
; Patent No. 6492131
; GENERAL INFORMATION:
; APPLICANT: Dieter Soll
; APPLICANT: Michael Ibba
; TITLE OF INVENTION: Class I-type Lysyl-tRNA Synthetase
; FILE REFERENCE: OCR-896
; CURRENT APPLICATION NUMBER: US/09/508,370A
; CURRENT FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: PCT/US98/18968
; PRIOR FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: MS Dos
; SEQ ID NO 3
; LENGTH: 521
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: lysyl t-RNA synthetase
; OTHER INFORMATION: construct expressed in Example 3
US-09-508-370A-3

Query Match          10.5%; Score 81; DB 4; Length 521;
Best Local Similarity 21.5%; Pred. No. 1.7;
Matches 32; Conservative 22; Mismatches 49; Indels 46; Gaps 4;

QY 14 BEQVELGFKGQIGTKDQYLEGKEYGVQTFQRFLLIIGYIOELMKFWLSHIDQYNNSSS 73
Db 345 KEEKKRAFK--RIYELSQPYMPSKEIPVQVGRHLVLSIQFE-----NNINK 390

QY 74 LRNHLNLEDDIMA-----QISITNGKDEVEDYEKNIKAKRNKL 111
Db 391 ILNLYLKNVQDQOKOLINKINCINAINWIRDFAPDKFSLRSKFDNMEILENSKAINEL 450

QY 112 -----RVIASITKETWKIDSLDL 130
Db 451 LDPLKKNFEVATEQDQIQNEIYKISRENNI 479

RESULT 4
US-08-392-625-17
; Sequence 17, Application US/08392625
; Patent No. 5837485
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
```

```
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5837485bert
; APPLICANT: Augustin, Johannes
; APPLICANT: Engelke, Germar
; APPLICANT: Rosenstein, Ralf
; APPLICANT: Kaletta, Cortina
; APPLICANT: Klein, Cora
; APPLICANT: Wieland, Bernd
; APPLICANT: Kupke, Thomas
; APPLICANT: Jung, G nther
; APPLICANT: Kellner, Roland
; TITLE OF INVENTION: Biosynthetic process For The Preparation
; TITLE OF INVENTION: Of Chemical Compounds
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,625
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,791
; FILING DATE: 30-APR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-392-625-17

Query Match          10.4%; Score 80.5; DB 2; Length 275;
Best Local Similarity 27.5%; Pred. No. 0.8;
Matches 36; Conservative 24; Mismatches 44; Indels 27; Gaps 8;

QY 21 FKEGQIQGTQDQYLEGKEYGVQTFQRFLLIIGYIO-ELMKFWLSHIDQ 67
Db 13 YKSFQSDLKXKLYINGT---YETNLGALINLITSGCGSEGINIYVGHLDLSISNIQ 69

QY 68 YNNSSSLRNHLNLEDDIMAQI--SITNGKDEVEDYEKNIKAKRNKLRLVIASITKETW--K 123
Db 70 KDKTFF--NLONIE--MKQVENVYTKITLVEEPEKNIKLIEDSSGIARQV--ETWYST 122

QY 124 IDSLDNLVKEV 134
Db 123 ISVIPNIIQEL 133

RESULT 5
US-08-466-961A-17
; Sequence 17, Application US/08466961A
; Patent No. 5843709
; GENERAL INFORMATION:
; APPLICANT: Entian, Karl-Dieter
; APPLICANT: G tz, Friedrich
; APPLICANT: Schnell, No. 5843709bert
; APPLICANT: Augustin, Johannes
```

RESULT 8
US-08-923-992A-4
; Sequence 4, Application US/08923992A
; Patent No. 6280738

```

; GENERAL INFORMATION:
; APPLICANT: Tai, Joseph Y.
; ADDRESSEE: Blake, Milan S.
; TITLE OF INVENTION: No. 6280738-IgA Fc Binding Forms of the Group B
; TITLE OF INVENTION: Streptococcal Beta Antigens
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,992A
; FILING DATE: 05-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,707
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 1438.0140001/RWE
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-923-992A-4

Query Match 10.1%; Score 78; DB 3; Length 1104;
Best Local Similarity 24.0%; Pred. No. 10;
Matches 36; Conservative 26; Mismatches 64; Indels 24; Gaps 5;

QY 5 DIDVNLNEEYELGFGKQIQTKDQYLGKEYGYGTGQRFLLIGYIOEL----- 57
DB 90 DVDNILSHEQKNEPKTKIDETNDS-DALLELEQFNET--NRLLIHQHEEVEKHNKP 146
QY 58 ----MKFWLSHIDQYNNSSSLRNHLNLEDDIMAQISITNGDKE-----VEDYEKNIKKAR 108
DB 147 QOKTLKQSDTKVDLSNIDKELNHQKSQVEAMAEQAQINEDKDSMLKKIEDIRKQAQAD 206
QY 109 NKLRVIASITKETWKIDS-----LDNLVKE 133
DB 207 KKEDAEVKVREELGKLFSTRAGLDQIQE 236

RESULT 9
US-08-867-611-10
; Sequence 10, Application US/08867611
; Patent No. 6172189
; GENERAL INFORMATION:
; APPLICANT: DEVARE, SUSHIL G
; APPLICANT: DESAI, SURESH M
; APPLICANT: CASEY, JAMES M
; APPLICANT: DAILEY, STEPHEN H
; APPLICANT: DAWSON, GEORGE J
; APPLICANT: GUTIERREZ, ROBIN A
; APPLICANT: LESNIEWSKI, RICHARD R
; APPLICANT: STEWART, JAMES L
; APPLICANT: RUPPRECHT, KEVIN R
; TITLE OF INVENTION: HEPATITIS C ASSAY UTILIZING RECOMBINANT
; TITLE OF INVENTION: ANTIGENS

```

```

; NUMBER OF SEQUENCES: 59
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: ONE ABBOTT PARK ROAD, CHAD377/AP6D2
; CITY: ABBOTT PARK
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,611
; FILING DATE: 02-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,757
; FILING DATE:
; APPLICATION NUMBER: US/08/179,896
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,822
; FILING DATE: 24-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/614,069
; FILING DATE: 07-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,561
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,565
; FILING DATE: 21-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/748,566
; FILING DATE: 21-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: FOREMSKI, PRISCILLA E
; REGISTRATION NUMBER: 33,207
; REFERENCE/DOCKET NUMBER: 4834.US.P6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708-937-6365
; TELEFAX: 708-937-9556
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 496 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-867-611-10

Query Match 10.0%; Score 77.5; DB 3; Length 496;
Best Local Similarity 20.4%; Pred. No. 3.9;
Matches 37; Conservative 22; Mismatches 67; Indels 55; Gaps 6;

QY 12 LEEEOYELGFGKQIQTKDQYLGKE-----YGYTGTFQRFLLI----- 50
DB 146 LDREGYALYFSRATIPWDRDRPAEGLETVDNLFHLGLGYGRAGEFIRYVNWQSPLEH 205
QY 51 IGYIQLMKFWLS-----HI-----DQYNNSSSLRNHLN----- 79
DB 206 IEMLEQLRVLTWYGEKIHVAQAEVPGTGYDTPEDLDPSNLSLDHYQDYLKEVKAASKV 265
QY 80 --NLEDIMAQISITNGDKEVEDEYKNIKARNLRVIASITKETWKIDSLDNLVKEVGQT 137
DB 266 KANLLSVEEACSLTPPHSAKSGFYGAKDVRCHARKAVTHINSVWK-DLLEDNVTPTDT 324
QY 138 L 138
DB 325 I 325

```

RESULT 11
US-08-500-222-2
; Sequence 2, Application US/08500222
; Patent No. 5643733
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:

RESULT 12
US-08-500-125-2
; Sequence 2, Application US/08500125
; Patent No. 5643751
; GENERAL INFORMATION:
; APPLICANT: Robinson, John M
; APPLICANT: Pilot-Matias, Tami J
; APPLICANT: Hunt, Jeffrey C
; TITLE OF INVENTION: Borrelia burgdorferi antigens
; TITLE OF INVENTION: and uses thereof
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA

APPLICATION NUMBER: US/08/867,611
FILING DATE: 02-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/646,757
FILING DATE:
APPLICATION NUMBER: US/08/179,896
FILING DATE:
APPLICATION DATA:
APPLICATION NUMBER: US 07/572,822
FILING DATE: 24-AUG-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/614,069
FILING DATE: 07-NOV-1990
APPLICATION DATA:
APPLICATION NUMBER: US 07/748,561
FILING DATE: 21-AUG-1991
APPLICATION DATA:
APPLICATION NUMBER: US 07/748,565
FILING DATE: 21-AUG-1991
APPLICATION DATA:
APPLICATION NUMBER: US 07/748,566
FILING DATE: 21-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834.US.P6
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-867-611-22

Query Match 10.0%; Score 77; DB 3; Length 425;
Best Local Similarity 23.1%; Pred. No. 3.5;
Matches 30; Conservative 19; Mismatches 43; Indels 38; Gaps 4;
QY 12 LEEQYELGFKGQIGTKQDYLEGKE-----YGVQTGFQRFLLI----- 50
Db 146 LDAEGYALYFSRATIPWDRDRFAEGLETVDGNFLRHLGIYGRAGFIRRYVNWQPSPLEH 205
QY 51 IGYIQELMKFWLSHIDQYNNSSSLRNHLNLEIDIMAQISITNGDKEVEDYEKN--IKKAR 108
Db 206 IEMLEQLRVLYWG-----EKIHVAVAEVPGTGVDTPELDLPSTRIRSR 250
QY 109 NKLRVIASIT 118
Db 251 NLGKVIDTLT 260

RESULT 15
PCT-US92-06965A-27
Sequence 27, Application PC/TUS9206965A
GENERAL INFORMATION:
APPLICANT: DEVARE, S.
APPLICANT: DESAI, S.
APPLICANT: DAILEY, S.
TITLE OF INVENTION: HCV SYNTHETIC PEPTIDE FROM NS1 REGION
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: ABBOTT LABORATORIES
STREET: ONE ABBOTT PARK ROAD
CITY: ABBOTT PARK
STATE: ILLINOIS
COUNTRY: U.S.
ZIP: 60065-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06965A
FILING DATE: 19920821
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: FOREMSKI, PRISCILLA E.
REGISTRATION NUMBER: 33,207
REFERENCE/DOCKET NUMBER: 4834PC.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-6365
TELEFAX: 708-937-9556
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 425 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-06965A-27

Query Match 10.0%; Score 77; DB 5; Length 425;
Best Local Similarity 23.1%; Pred. No. 3.5;
Matches 30; Conservative 19; Mismatches 43; Indels 38; Gaps 4;
QY 12 LEEQYELGFKGQIGTKQDYLEGKE-----YGVQTGFQRFLLI----- 50
Db 146 LDAEGYALYFSRATIPWDRDRFAEGLETVDGNFLRHLGIYGRAGFIRRYVNWQPSPLEH 205
QY 51 IGYIQELMKFWLSHIDQYNNSSSLRNHLNLEIDIMAQISITNGDKEVEDYEKN--IKKAR 108
Db 206 IEMLEQLRVLYWG-----EKIHVAVAEVPGTGVDTPELDLPSTRIRSR 250
QY 109 NKLRVIASIT 118
Db 251 NLGKVIDTLT 260

Search completed: February 16, 2004, 14:15:31
Job time : 31 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2004, 12:53:01 ; Search time 39 Seconds
(without alignments)
602.347 Million cell updates/sec

Title: US-09-980-054A-12

Perfect score: 772

Sequence: 1 MSDDIDNVLNLEEEQYELG.....NLVKEVGGLQVSENPDMMW 148

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq.19Jun03.*

```

1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	772	100.0	148	22 AAB30859	Amino acid sequenc
2	764	99.0	148	22 AAU15103	Protein encoded by
3	764	99.0	148	23 ABP73225	Candida albicans e
4	116	15.0	404	21 AAG30345	Arabidopsis thalia
5	99.5	12.9	237	22 ABG28105	Novel human diago
6	99.5	12.9	293	22 ABB11923	Human GK003 homolo
7	99.5	12.9	293	22 AAM79451	Human protein SEQ
8	99.5	12.9	310	22 AAM78467	Human protein SEQ
9	99	12.8	107	21 AAG01966	Human secreted pro

10	95.5	12.4	141	22 ABB67857	Drosophila melanog
11	95	12.3	1646	23 ABP69013	Human polypeptide
12	90.5	11.7	525	23 AAU12041	Clostridium diffic
13	90	11.7	475	22 ABB68543	Drosophila melanog
14	87.5	11.3	803	23 AAU12040	Clostridium diffic
15	87	11.3	239	20 AAY20137	B. burgdorferi ant
16	87	11.3	274	20 AAY20136	B. burgdorferi ant
17	86	11.1	2285	20 AAW98149	Bacillus subtilis
18	85	11.0	507	21 AAB18173	Plasmodium falcipa
19	83.5	10.8	880	22 AAB96332	Putative P. abyssi
20	83.5	10.8	930	19 AAW55961	Human transient re
21	83.5	10.8	930	24 ABB99469	Amino acid sequenc
22	83.5	10.8	1211	22 ABB64663	Drosophila melanog
23	83	10.8	1184	22 AAU33733	Staphylococcus aur
24	83	10.8	1188	22 AAU36532	Staphylococcus aur
25	82.5	10.7	185	21 AAG54838	Arabidopsis thalia
26	82.5	10.7	435	23 ABP73423	Candida albicans e
27	82	10.6	724	22 AAG82994	S. epidermidis ope
28	82	10.6	756	23 ABP38708	Staphylococcus epi
29	82	10.6	1516	21 AAB18195	Plasmodium falcipa
30	81	10.5	320	22 AAU38192	Salmonella typhi c
31	81	10.5	440	22 AAU47019	Propionibacterium
32	81	10.5	470	23 ABP30175	Streptococcus poly
33	81	10.5	476	23 ABP26844	Streptococcus poly
34	81	10.5	521	20 AAY04366	Borrelia burgdorfe
35	81	10.5	699	22 ABB64367	Drosophila melanog
36	81	10.5	1184	19 AAW40541	Mutant C-beta prot
37	81	10.5	1188	21 AAB18183	Plasmodium falcipa
38	80.5	10.4	254	23 ABP73674	Candida albicans e
39	80	10.4	280	22 ABB58792	Drosophila melanog
40	80	10.4	1156	20 AAY34753	Amino acid sequenc
41	79.5	10.3	368	23 ABU52170	Helicobacter pylor
42	79.5	10.3	600	19 AAW98568	H. pylori GHP0.504
43	79.5	10.3	600	22 ABB68337	Amino acid sequenc
44	79.5	10.3	600	22 ABB68338	Amino acid sequenc
45	79.5	10.3	677	22 AAB68339	Amino acid sequenc

ALIGNMENTS

RESULT 1

AAAB30859
ID AAB30859 standard; Protein; 148 AA.

XX AC AAB30859;

XX DT 02-APR-2001 (first entry)

XX DE Amino acid sequence of a Candida albicans protein CanL260.

XX XX CaDRA472; CaDRA489; 1CaDR527; 2CaDR527; CaPLO24; CanL260; CaDR361;
KW antifungal; fungal infection; pathogenic fungi.

XX OS Candida albicans.

XX XX WO2000075305-A2.

XX XX 14-DEC-2000.

XX PF 08-JUN-2000; 2000MO-FR01567.

XX XX 09-JUN-1999; 99FR-0007250.

XX XX (HMRI) HOECHST MARION ROUSSEL.

XX XX Lalanne J, Rocher C;

XX XX WPI; 2001-050024/06.

XX DR N-PSDB; AAC86750.

XX XX New polynucleotides from Candida albicans and their derived proteins,
PT useful for diagnosis and treatment of fungal infections and for drug

PT screening -
 XX Claim 11; Page 78-79; 89pp; French.
 PS
 CC The present sequence represents a Candida albicans protein. The
 CC specification describes genes CaDRA472, CaDR489, CaDR527, 2CaDR527,
 CC CaFIO24, CaN1260, and CaDR361. These genes are essential for
 CC survival, and so are good targets for antifungal agents. The Candida
 CC albicans genes and their derived proteins are used to screen compounds
 CC for the ability to inhibit the activity of the protein, i.e. for
 CC antifungal activity. The proteins are also used to generate a protective
 CC antibody response against fungal infections in mammals. The Candida
 CC albicans proteins and genes, and their antibodies, are used for
 CC diagnosing fungal infections, specifically C. albicans (in standard
 CC amplification, hybridisation or immunological assays, and for studying
 CC pathogenic fungi.
 XX
 SQ Sequence 148 AA;
 Query Match 100.0%; Score 772; DB 22; Length 148;
 Best Local Similarity 100.0%; Pred. No. 2.3e-71;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSDDIDNVNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 60
 DB 1 MSDDIDNVNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 60
 QY 61 WLSHIDQYNNSSSLRNHLNLEEDIMAOISITNGDKEVEDYEKNIKARNKLRVIASITKE 120
 DB 61 WLSHIDQYNNSSSLRNHLNLEEDIMAOISITNGDKEVEDYEKNIKARNKLRVIASITKE 120
 QY 121 TWKIDS LNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 148
 DB 121 TWKIDS LNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 148
 RESULT 2
 ID AAU15103
 AC AAU15103;
 DT 04-DEC-2001 (first entry)
 XX Protein encoded by C. albicans essential gene CaYNL260C.
 DE Gene identification; essential gene; GRACE; pathogenic fungus;
 KW gene replacement and conditional expression; fungal infection.
 XX Candida albicans.
 OS
 XX WO200160975-A2.
 PN 23-AUG-2001.
 XX 20-FEB-2001; 2001WO-US05551.
 XX 18-FEB-2000; 2000US-0183534.
 PR (ELIT-) ELITRA PHARM INC.
 PA Roemer T, Jiang B, Boone C, Bussey H;
 PI WPI; 2001-489080/53.
 DR N-PSDB; AAS23431.
 XX Identifying genes essential to fungal metabolisms and identifying
 PT potential therapeutic agents that target these genes -
 XX Claim 43; Page 237-238; 324pp; English.
 PS The present invention relates to novel methods for constructing fungal
 XX strains useful for identification and validation of gene products as

CC targets for therapeutic agents, for creating a collection of identified
 CC essential genes, and screening assays for the discovery of new drugs.
 CC The invention provides the GRACE (gene replacement and conditional
 CC expression) method for the construction of mutant organisms referred to
 CC as GRACE strains of the organism. The invention can be applied to any
 CC organism, particularly a pathogenic fungus e.g. Candida albicans,
 CC Aspergillus fumigatus and Cryptococcus neoformans. The methods are
 CC useful to identify agents that may be used in the treatment of fungal
 CC infections. AAU15051-AAU15113 represent proteins encoded by C. albicans
 CC essential genes.
 XX
 SQ Sequence 148 AA;
 Query Match 99.0%; Score 764; DB 22; Length 148;
 Best Local Similarity 98.6%; Pred. No. 1.5e-70;
 Matches 146; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSDDIDNVNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 60
 DB 1 MSDDIDNVNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 60
 QY 61 WLSHIDQYNNSSSLRNHLNLEEDIMAOISITNGDKEVEDYEKNIKARNKLRVIASITKE 120
 DB 61 WLSHIDQYNNSSSLRNHLNLEEDIMAOISITNGDKEVEDYEKNIKARNKLRVIASITKE 120
 QY 121 TWKIDS LNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 148
 DB 121 TWKIDS LNLNLEEEQYELGFKGQIQGTDQYLEGKEYGYQTGFQRFLLIIGYIQELMKF 148
 RESULT 3
 ID ABP73225
 AC ABP73225;
 DT 30-JAN-2003 (first entry)
 XX Candida albicans essential protein SEQ ID NO 7062.
 DE Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal.
 XX Candida albicans.
 OS
 XX WO200253728-A2.
 PN 11-JUL-2002.
 XX 26-DEC-2001; 2001WO-US49486.
 XX 29-DEC-2000; 2000US-359128P.
 PR 20-FEB-2001; 2001US-0792024.
 PR 22-AUG-2001; 2001US-314050P.
 XX (ELIT-) ELITRA PHARM INC.
 PA Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 PI WPI; 2002-566694/60.
 DR N-PSDB; ABZ31775.
 XX Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele
 PT of a gene and placing other allele of the gene under conditional
 PT expression -
 XX Claim 44; SEQ ID NO 7062; 167pp + Sequence Listing; English.
 PS The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an

CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of *C. albicans* cells and for
 CC treating infection by *C. albicans*. The present sequence is that of an
 CC essential *Candida albicans* protein used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.

XX SQ Sequence 148 AA;

Query Match 99.0%; Score 764; DB 23; Length 148;
 Best Local Similarity 98.6%; Pred. No. 1.5e-70;
 Matches 146; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSDDIDNVLNLEEQYELGPKGQIQGKQYLGKGYGTGQFRLIIGYIOELMKF 60
 DB 1 MSDDIDNVLNLEEQYELGPKGQIQGKQYLGKGYGTGQFRLIIGYIOELMKF 60
 QY 61 WLSHIDQYNNSSLRNHNLEDDMAQISITNGKVEDYKNTKKARKLRVATISITKE 120
 DB 61 WLSHIDQYNNSSLRNHNLEDDMAQISITNGKVEDYKNTKKARKLRVATISITKE 120
 QY 121 TWKIDSLNVLKVEGGTQLQVSNPDPMW 148
 DB 121 TWKIDSLNVLKVEGGTQLQVSNPDPMW 148

RESULT 4

AAG30345

ID AAG30345 standard; Protein; 404 AA.

XX AC AAG30345;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 36261.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 XX hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 05-MAY-1999; 99US-0132486.
 PR 05-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 18-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 23-JUN-1999; 99US-0140931.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 02-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 05-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 08-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ_ID No 58464; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (II) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 237 AA;
Query Match 12.9%; Score 99.5; DB 22; Length 237;
Best Local Similarity 26.7%; Pred. No. 0.053;
Matches 32; Conservative 23; Mismatches 46; Indels 19; Gaps 3;
QY 11 NLEEEYELGKEGQIQTKQOYLEGKEYGQTGFQFLIIGYQELMKFWLSHIDQYNN 70
DB 121 NMQRVKE-GYRDGIDAGKAVTLQGGFNQGYKGAELVILNYGRILGTSALLSWCHLNN 179
QY 71 SSSLRNHLNLEDIMAQISITNGDKVEDYENKIKKARKNLKVIASTKTKWIDSLDNL 130
DB 180 NSTLINKINLLDVGQ-----CEY-----VLKHLKSIPTPSHVVDLDSI 221
RESULT 6
ID ABB11923
AC ABB11923;
XX
XX 11-JAN-2002 (first entry)
DE Human GK003 homologue, SEQ ID NO:2293.
KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antischmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytopathic; osteopathic; vasoropitic; cardiant; virucide; antibacterial;
KW antifungal; vulnery; antiulcer.
XX
OS Homo sapiens.
XX
XX WO200157198-A2.
XX

PD 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US03800.
XX
XX 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
XX (HYSB-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-457740/49.
XX
XX N-PSDB; ABA09167.
DR
DR Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
XX Claim 20; Page 277; 1963pp; English.
XX
CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC invention also relates to vectors and recombinant host cells comprising a
CC nucleotide of the invention, methods of producing the novel polypeptides,
CC antibodies against the polypeptides, methods of detecting the nucleotides
CC or polypeptides in a sample, and methods of identifying compounds which
CC bind to polypeptides of the invention. Although novel many of the
CC polypeptides of the invention have homology to known proteins, thereby
CC giving an insight into their probable biological activities, and hence
CC potential therapeutic applications. The polypeptides of the invention may
CC have various activities, including cytokine, cell proliferation or cell
CC differentiation activities; stem cell growth factor activity;
CC haematopoiesis regulatory activity; tissue growth activity;
CC immunomodulatory activity; activin- or inhibin-related activities;
CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC thrombolytic activities; receptor or ligand activities; or may be
CC involved in oncogenesis, cancer cell proliferation or metastasis.
CC Depending on their biological activities, polypeptides and nucleotides of
CC the invention are useful for preventing, treating or ameliorating medical
CC conditions, e.g., by protein or gene therapy. Such conditions include
CC cancers, haematopoietic disorders (e.g. myeloid or lymphoid cell
CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC proliferative retinopathy, atherosclerosis, coronary heart disease,
CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC vascular growth. Polypeptides involved with tissue regeneration and
CC repair (or nucleic acids encoding them) may be used to promote wound
CC healing (e.g., of burns, incisions and ulcers), while those with
CC immunomodulatory activities may be used in the treatment of viral,
CC bacterial and fungal infections in addition to immune disorders.
CC Polypeptides with growth factor activity may be used in cell cultures to
CC promote cell growth. For example, such polypeptides may be used to
CC manipulate stem cells in culture to give rise to neuroepithelial cells
CC that can be used to augment or replace cells damaged by illness,
CC autoimmune disease or accidental damage. The polypeptides and nucleotides
CC may also be used in the diagnosis of the above conditions, and in drug
CC screening techniques. The present sequence represents a novel human
CC polypeptide of the invention.
XX
SQ Sequence 293 AA;
Query Match 12.9%; Score 99.5; DB 22; Length 293;
Best Local Similarity 26.7%; Pred. No. 0.07;
Matches 32; Conservative 23; Mismatches 46; Indels 19; Gaps 3;
QY 11 NLEEEYELGKEGQIQTKQOYLEGKEYGQTGFQFLIIGYQELMKFWLSHIDQYNN 70
DB 121 NMQRVKE-GYRDGIDAGKAVTLQGGFNQGYKGAELVILNYGRILGTSALLSWCHLNN 179
QY 71 SSSLRNHLNLEDIMAQISITNGDKVEDYENKIKKARKNLKVIASTKTKWIDSLDNL 130
DB 180 NSTLINKINLLDVGQ-----CEY-----VLKHLKSIPTPSHVVDLDSI 221

QY 71 SSSLRNHLNLEDDMAQISTNGKXEDYENKIKKARNKLRVIASTKETWKIDSLNL 130
 Db 180 NSTLINKINLLDAVQ-----CEY-----VLKHLKSITPPSHVVDLDSI 221

RESULT 9
 AAG01966
 ID AAG01966 standard; Protein; 107 AA.

AC AAG01966;
 XX
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein, SEQ ID NO: 6047.

DE Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping.
 XX Homo sapiens.

OS
 XX
 PN EP1033401-A2.
 XX
 PD 06-SEP-2000.

XX
 PF 21-FEB-2000; 2000BP-0200610.
 XX
 PR 26-FEB-1999; 99US-0122487.

XX
 PA (GEST) GENSET.
 XX
 PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX
 DR WPI; 2000-500381/45.
 DR N-PSDB; AAC01972.

XX
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 XX
 PS Claim 13; SEQ ID 6047; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30
 CC different tissues. EST sequences usually correspond mainly to the 3'
 CC untranslated region (UTR) of the mRNA because they are often obtained
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
 CC those cases where longer cDNA sequences have been obtained, the full 5'
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
 CC ends and can therefore be used to obtain full length cDNAs and genomic
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
 CC chromosome mapping procedures. They are used to obtain upstream
 CC regulatory sequences and to design expression and secretion vectors.

XX Sequence 107 AA;

Query Match 12.8%; Score 99; DB 21; Length 107;
 Best Local Similarity 27.0%; Pred. No. 0.021;
 Matches 30; Conservative 21; Mismatches 42; Indels 18; Gaps 2;

QY 20 GKKEGQIQTKQYLEGKGYGTGFORFLIIGYIOELMKFWLSHIDVNNSSSLRNHLN 79
 Db 8 GYRDGIDAGKAVTLQGFNGQYKKAQVILNYGRLEGILSALLSWCHLNNSTLINKIN 67

QY 80 NLEDIMAQISTNGKXEDYENKIKKARNKLRVIASTKETWKIDSLNL 130
 Db 68 NLLDAVQ-----CEY-----VLKHLKSITPPSHVVDLDSI 100

RESULT 10
 ABB67857

ID ABB67857 standard; Protein; 141 AA.

AC ABB67857;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 30363.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.

XX Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

PR 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

PA (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11960.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -

XX Disclosure; SEQ ID NO 30363; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (AB57737-AB572072).

CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 141 AA;

Query Match 12.4%; Score 95.5; DB 22; Length 141;
 Best Local Similarity 25.5%; Pred. No. 0.068;
 Matches 28; Conservative 29; Mismatches 42; Indels 11; Gaps 5;

QY 7 DNVNLLEBQYELGFKKGQIQGTQYLEGKGYGTGFORFLIIGYIOELMKFWLSHI - 65
 Db 13 DDIVLTKEARLGYEE---GLKDGQEQNGRGGYKGLGVAQVSLG--BELGKI-LQGVV 65

QY 66 --DYNSSSLRNHLNLEDDIMAQISTNGDK-EVEDYENKIKKARNKLR 112

Db 66 AQOQLKHTDKVRSLRSLRSLIEFFRTNDPQADIVGAVQDIRSHRRLR 115

RESULT 11

ABP69013

ID ABP69013 standard; Protein; 1646 AA.

XX ABP69013;

AC ABP69013;

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1060.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;

KW cell-proliferative disorder; neurodegenerative disease; bacterial;
 KW Parkinson's disease; Alzheimer's disease; autoimmune disease;
 KW multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
 KW arthritis; cytostatic; immunomodulator; neutropenic; neuroprotective;
 KW antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
 KW haemostatic; vulvular; fungicide; antibacterial; virucide; protozoacide;
 KW antiarthritic.
 XX Homo sapiens.
 OS
 XX WO200270539-A2.
 PN
 XX 12-SEP-2002.
 PD
 XX
 XX 05-MAR-2002; 2002WO-US05095.
 XX
 XX 05-MAR-2001; 2001US-0799451.
 PR
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren P;
 PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
 PI Wehrman T, Wang J, Wang D, Drmanac R;
 XX WPI; 2002-759812/82.
 DR N-PSDB; ABZ11230.
 DR
 XX New polynucleotides comprising sequences assembled from expressed
 PT sequence tags (ESTs), useful for treating cell-proliferative,
 PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or
 PT platelet or coagulation disorders -
 XX Claim 9; SEQ ID NO 1060; 1012pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence selected from any of 948 sequences
 CC (ABZ11119-ABZ12056) or their mature protein coding portion, active domain
 CC coding protein or complementary sequences. The polynucleotides are useful
 CC for identifying expressed genes or for physical mapping of human genome.
 CC The encoded polypeptides (ABP68902-ABP69849) are useful as molecular
 CC weight markers, as a food supplement, for generating antibodies, in
 CC medical imaging, screening and diagnostic assays and for treating
 CC cell-proliferative disorders (cancer), neurodegenerative diseases
 CC (Parkinson's or Alzheimer's disease), autoimmune diseases (multiple
 CC sclerosis, diabetes, lupus) genetic disorders, myeloid or lymphoid
 CC disorders, platelet or coagulation disorders, wound, burns, incision,
 CC ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,
 CC parasitic), arthritis, etc.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1646 AA;
 Query Match 12.3%; Score 95; DB 23; Length 1646;
 Best Local Similarity 25.2%; Pred. No. 2.1;
 Matches 36; Conservative 27; Mismatches 40; Indels 40; Gaps 7;
 QY 24 QIQGTQKQYLEGKEYGY-----QTGFQRFLLIIGYIQELMKFWLSH----- 64
 DB 626 GALKQKVEDELELDKSFETLADQTEWQSHLFYQFEWQLWEAHSSELLVQLELEKR 685
 QY 65 IDVNNSSSLRN-----HLNLEDIMAQISITNGDKE-----VEDYENKIKKARNKL 111
 DB 686 MEQHQKHSLSQVQEAHLDRLLQLRQS-----DKETLAFLEKVKDKLQKMSRYECF 741
 QY 112 RVIASITKWTWKIDSLDLNIVKEV 134
 DB 742 HTL--LTKVEVMEYPAI--MLKEL 760
 RESULT 12
 AAU12041

ID AAU12041 standard; Protein; 525 AA.
 XX
 AC AAU12041;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Clostridium difficile S-layer protein cell wall binding portion (ORF7).
 XX
 KW Gene expression cassette; secretory leader sequence; anti-tumour therapy;
 KW clostridial N-acetylmuramoyl-L-alanine amidase-like protein; cytostatic;
 KW fibrinolysis; gangrenous tissue; necrotic tissue; infection;
 KW S-layer protein cell wall binding portion; ORF7.
 XX
 OS Clostridium difficile.
 XX
 PN WO200194599-A1.
 PD 13-DEC-2001.
 XX
 XX 07-JUN-2001; 2001WO-SE01280.
 XX
 XX 07-JUN-2000; 2000SE-0002139.
 PR 26-APR-2001; 2001SE-0101479.
 XX
 XX (SMIT-) SMITTSKYDDSDINSTITUTET.
 PA
 XX Burman LG, Akerlund T, Mukherjee K, Katagihallimath N;
 PI WPI; 2002-122121/16.
 DR
 XX New gene expression cassette comprising a secretory leader sequence
 PT encoding a signal peptide from Clostridium difficile, useful e.g. for
 PT presenting polypeptides on bacterial cell surface, or as anti-tumour
 PT therapy -
 XX Claim 6; Page 77-79; 113pp; English.
 PS
 XX The present invention relates to a new gene expression cassette
 CC comprising a secretory leader sequence encoding a signal peptide from
 CC Clostridium difficile or signal peptides of analogous exported
 CC clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked
 CC to a DNA sequence encoding a heterologous polypeptide. The gene
 CC expression cassette is useful for presenting polypeptides on the surface
 CC of bacterial cells, and/or secreting them into the surroundings of the
 CC cells, for mucosal immunisation, induction of immunological tolerance
 CC and anti-tumour therapy in humans and animals, as vaccines, and for the
 CC treatment of prevention of fibrinolysis in arterial or venous occlusion
 CC and/or revitalising gangrenous or necrotic tissue in various diseases.
 CC The gene expression cassette may also be used to produce in the gut,
 CC e.g. peptides and enzymes for therapy and prophylaxis of various
 CC diseases; single, fusion or multiple polypeptide antigens of microbial,
 CC animal or mammalian origin for neonatal immune balancing, vaccination
 CC against infections; and carrier molecules separated or fused to the
 CC antigen to amplify or modulate immune response. Spores produced by
 CC Clostridia, e.g. Clostridium difficile are useful for both industrial
 CC production of vaccines and for local production of the desired
 CC polypeptides at the body sites desired. AAU12037-AAU12048 represent
 CC C. difficile S-layer protein cell wall binding portions encoded by
 CC ORF1, ORF3, ORF5-7, ORF9, ORF11, ORF12, ORF13, ORF14, ORF15, ORF16, ORF17, ORF18, ORF19, ORF20, ORF21, ORF22, ORF23, ORF24, ORF25, ORF26, ORF27, ORF28, ORF29, ORF30, ORF31, ORF32, ORF33, ORF34, ORF35, ORF36, ORF37, ORF38, ORF39, ORF40, ORF41, ORF42, ORF43, ORF44, ORF45, ORF46, ORF47, ORF48, ORF49, ORF50, ORF51, ORF52, ORF53, ORF54, ORF55, ORF56, ORF57, ORF58, ORF59, ORF60, ORF61, ORF62, ORF63, ORF64, ORF65, ORF66, ORF67, ORF68, ORF69, ORF70, ORF71, ORF72, ORF73, ORF74, ORF75, ORF76, ORF77, ORF78, ORF79, ORF80, ORF81, ORF82, ORF83, ORF84, ORF85, ORF86, ORF87, ORF88, ORF89, ORF90, ORF91, ORF92, ORF93, ORF94, ORF95, ORF96, ORF97, ORF98, ORF99, ORF100, ORF101, ORF102, ORF103, ORF104, ORF105, ORF106, ORF107, ORF108, ORF109, ORF110, ORF111, ORF112, ORF113, ORF114, ORF115, ORF116, ORF117, ORF118, ORF119, ORF120, ORF121, ORF122, ORF123, ORF124, ORF125, ORF126, ORF127, ORF128, ORF129, ORF130, ORF131, ORF132, ORF133, ORF134, ORF135, ORF136, ORF137, ORF138, ORF139, ORF140, ORF141, ORF142, ORF143, ORF144, ORF145, ORF146, ORF147, ORF148, ORF149, ORF150, ORF151, ORF152, ORF153, ORF154, ORF155, ORF156, ORF157, ORF158, ORF159, ORF160, ORF161, ORF162, ORF163, ORF164, ORF165, ORF166, ORF167, ORF168, ORF169, ORF170, ORF171, ORF172, ORF173, ORF174, ORF175, ORF176, ORF177, ORF178, ORF179, ORF180, ORF181, ORF182, ORF183, ORF184, ORF185, ORF186, ORF187, ORF188, ORF189, ORF190, ORF191, ORF192, ORF193, ORF194, ORF195, ORF196, ORF197, ORF198, ORF199, ORF200, ORF201, ORF202, ORF203, ORF204, ORF205, ORF206, ORF207, ORF208, ORF209, ORF210, ORF211, ORF212, ORF213, ORF214, ORF215, ORF216, ORF217, ORF218, ORF219, ORF220, ORF221, ORF222, ORF223, ORF224, ORF225, ORF226, ORF227, ORF228, ORF229, ORF230, ORF231, ORF232, ORF233, ORF234, ORF235, ORF236, ORF237, ORF238, ORF239, ORF240, ORF241, ORF242, ORF243, ORF244, ORF245, ORF246, ORF247, ORF248, ORF249, ORF250, ORF251, ORF252, ORF253, ORF254, ORF255, ORF256, ORF257, ORF258, ORF259, ORF260, ORF261, ORF262, ORF263, ORF264, ORF265, ORF266, ORF267, ORF268, ORF269, ORF270, ORF271, ORF272, ORF273, ORF274, ORF275, ORF276, ORF277, ORF278, ORF279, ORF280, ORF281, ORF282, ORF283, ORF284, ORF285, ORF286, ORF287, ORF288, ORF289, ORF290, ORF291, ORF292, ORF293, ORF294, ORF295, ORF296, ORF297, ORF298, ORF299, ORF300, ORF301, ORF302, ORF303, ORF304, ORF305, ORF306, ORF307, ORF308, ORF309, ORF310, ORF311, ORF312, ORF313, ORF314, ORF315, ORF316, ORF317, ORF318, ORF319, ORF320, ORF321, ORF322, ORF323, ORF324, ORF325, ORF326, ORF327, ORF328, ORF329, ORF330, ORF331, ORF332, ORF333, ORF334, ORF335, ORF336, ORF337, ORF338, ORF339, ORF340, ORF341, ORF342, ORF343, ORF344, ORF345, ORF346, ORF347, ORF348, ORF349, ORF350, ORF351, ORF352, ORF353, ORF354, ORF355, ORF356, ORF357, ORF358, ORF359, ORF360, ORF361, ORF362, ORF363, ORF364, ORF365, ORF366, ORF367, ORF368, ORF369, ORF370, ORF371, ORF372, ORF373, ORF374, ORF375, ORF376, ORF377, ORF378, ORF379, ORF380, ORF381, ORF382, ORF383, ORF384, ORF385, ORF386, ORF387, ORF388, ORF389, ORF390, ORF391, ORF392, ORF393, ORF394, ORF395, ORF396, ORF397, ORF398, ORF399, ORF400, ORF401, ORF402, ORF403, ORF404, ORF405, ORF406, ORF407, ORF408, ORF409, ORF410, ORF411, ORF412, ORF413, ORF414, ORF415, ORF416, ORF417, ORF418, ORF419, ORF420, ORF421, ORF422, ORF423, ORF424, ORF425, ORF426, ORF427, ORF428, ORF429, ORF430, ORF431, ORF432, ORF433, ORF434, ORF435, ORF436, ORF437, ORF438, ORF439, ORF440, ORF441, ORF442, ORF443, ORF444, ORF445, ORF446, ORF447, ORF448, ORF449, ORF450, ORF451, ORF452, ORF453, ORF454, ORF455, ORF456, ORF457, ORF458, ORF459, ORF460, ORF461, ORF462, ORF463, ORF464, ORF465, ORF466, ORF467, ORF468, ORF469, ORF470, ORF471, ORF472, ORF473, ORF474, ORF475, ORF476, ORF477, ORF478, ORF479, ORF480, ORF481, ORF482, ORF483, ORF484, ORF485, ORF486, ORF487, ORF488, ORF489, ORF490, ORF491, ORF492, ORF493, ORF494, ORF495, ORF496, ORF497, ORF498, ORF499, ORF500, ORF501, ORF502, ORF503, ORF504, ORF505, ORF506, ORF507, ORF508, ORF509, ORF510, ORF511, ORF512, ORF513, ORF514, ORF515, ORF516, ORF517, ORF518, ORF519, ORF520, ORF521, ORF522, ORF523, ORF524, ORF525, ORF526, ORF527, ORF528, ORF529, ORF530, ORF531, ORF532, ORF533, ORF534, ORF535, ORF536, ORF537, ORF538, ORF539, ORF540, ORF541, ORF542, ORF543, ORF544, ORF545, ORF546, ORF547, ORF548, ORF549, ORF550, ORF551, ORF552, ORF553, ORF554, ORF555, ORF556, ORF557, ORF558, ORF559, ORF560, ORF561, ORF562, ORF563, ORF564, ORF565, ORF566, ORF567, ORF568, ORF569, ORF570, ORF571, ORF572, ORF573, ORF574, ORF575, ORF576, ORF577, ORF578, ORF579, ORF580, ORF581, ORF582, ORF583, ORF584, ORF585, ORF586, ORF587, ORF588, ORF589, ORF590, ORF591, ORF592, ORF593, ORF594, ORF595, ORF596, ORF597, ORF598, ORF599, ORF600, ORF601, ORF602, ORF603, ORF604, ORF605, ORF606, ORF607, ORF608, ORF609, ORF610, ORF611, ORF612, ORF613, ORF614, ORF615, ORF616, ORF617, ORF618, ORF619, ORF620, ORF621, ORF622, ORF623, ORF624, ORF625, ORF626, ORF627, ORF628, ORF629, ORF630, ORF631, ORF632, ORF633, ORF634, ORF635, ORF636, ORF637, ORF638, ORF639, ORF640, ORF641, ORF642, ORF643, ORF644, ORF645, ORF646, ORF647, ORF648, ORF649, ORF650, ORF651, ORF652, ORF653, ORF654, ORF655, ORF656, ORF657, ORF658, ORF659, ORF660, ORF661, ORF662, ORF663, ORF664, ORF665, ORF666, ORF667, ORF668, ORF669, ORF670, ORF671, ORF672, ORF673, ORF674, ORF675, ORF676, ORF677, ORF678, ORF679, ORF680, ORF681, ORF682, ORF683, ORF684, ORF685, ORF686, ORF687, ORF688, ORF689, ORF690, ORF691, ORF692, ORF693, ORF694, ORF695, ORF696, ORF697, ORF698, ORF699, ORF700, ORF701, ORF702, ORF703, ORF704, ORF705, ORF706, ORF707, ORF708, ORF709, ORF710, ORF711, ORF712, ORF713, ORF714, ORF715, ORF716, ORF717, ORF718, ORF719, ORF720, ORF721, ORF722, ORF723, ORF724, ORF725, ORF726, ORF727, ORF728, ORF729, ORF730, ORF731, ORF732, ORF733, ORF734, ORF735, ORF736, ORF737, ORF738, ORF739, ORF740, ORF741, ORF742, ORF743, ORF744, ORF745, ORF746, ORF747, ORF748, ORF749, ORF750, ORF751, ORF752, ORF753, ORF754, ORF755, ORF756, ORF757, ORF758, ORF759, ORF760, ORF761, ORF762, ORF763, ORF764, ORF765, ORF766, ORF767, ORF768, ORF769, ORF770, ORF771, ORF772, ORF773, ORF774, ORF775, ORF776, ORF777, ORF778, ORF779, ORF780, ORF781, ORF782, ORF783, ORF784, ORF785, ORF786, ORF787, ORF788, ORF789, ORF790, ORF791, ORF792, ORF793, ORF794, ORF795, ORF796, ORF797, ORF798, ORF799, ORF800, ORF801, ORF802, ORF803, ORF804, ORF805, ORF806, ORF807, ORF808, ORF809, ORF810, ORF811, ORF812, ORF813, ORF814, ORF815, ORF816, ORF817, ORF818, ORF819, ORF820, ORF821, ORF822, ORF823, ORF824, ORF825, ORF826, ORF827, ORF828, ORF829, ORF830, ORF831, ORF832, ORF833, ORF834, ORF835, ORF836, ORF837, ORF838, ORF839, ORF840, ORF841, ORF842, ORF843, ORF844, ORF845, ORF846, ORF847, ORF848, ORF849, ORF850, ORF851, ORF852, ORF853, ORF854, ORF855, ORF856, ORF857, ORF858, ORF859, ORF860, ORF861, ORF862, ORF863, ORF864, ORF865, ORF866, ORF867, ORF868, ORF869, ORF870, ORF871, ORF872, ORF873, ORF874, ORF875, ORF876, ORF877, ORF878, ORF879, ORF880, ORF881, ORF882, ORF883, ORF884, ORF885, ORF886, ORF887, ORF888, ORF889, ORF890, ORF891, ORF892, ORF893, ORF894, ORF895, ORF896, ORF897, ORF898, ORF899, ORF900, ORF901, ORF902, ORF903, ORF904, ORF905, ORF906, ORF907, ORF908, ORF909, ORF910, ORF911, ORF912, ORF913, ORF914, ORF915, ORF916, ORF917, ORF918, ORF919, ORF920, ORF921, ORF922, ORF923, ORF924, ORF925, ORF926, ORF927, ORF928, ORF929, ORF930, ORF931, ORF932, ORF933, ORF934, ORF935, ORF936, ORF937, ORF938, ORF939, ORF940, ORF941, ORF942, ORF943, ORF944, ORF945, ORF946, ORF947, ORF948, ORF949, ORF950, ORF951, ORF952, ORF953, ORF954, ORF955, ORF956, ORF957, ORF958, ORF959, ORF960, ORF961, ORF962, ORF963, ORF964, ORF965, ORF966, ORF967, ORF968, ORF969, ORF970, ORF971, ORF972, ORF973, ORF974, ORF975, ORF976, ORF977, ORF978, ORF979, ORF980, ORF981, ORF982, ORF983, ORF984, ORF985, ORF986, ORF987, ORF988, ORF989, ORF990, ORF991, ORF992, ORF993, ORF994, ORF995, ORF996, ORF997, ORF998, ORF999, ORF1000, ORF1001, ORF1002, ORF1003, ORF1004, ORF1005, ORF1006, ORF1007, ORF1008, ORF1009, ORF1010, ORF1011, ORF1012, ORF1013, ORF1014, ORF1015, ORF1016, ORF1017, ORF1018, ORF1019, ORF1020, ORF1021, ORF1022, ORF1023, ORF1024, ORF1025, ORF1026, ORF1027, ORF1028, ORF1029, ORF1030, ORF1031, ORF1032, ORF1033, ORF1034, ORF1035, ORF1036, ORF1037, ORF1038, ORF1039, ORF1040, ORF1041, ORF1042, ORF1043, ORF1044, ORF1045, ORF1046, ORF1047, ORF1048, ORF1049, ORF1050, ORF1051, ORF1052, ORF1053, ORF1054, ORF1055, ORF1056, ORF1057, ORF1058, ORF1059, ORF1060, ORF1061, ORF1062, ORF1063, ORF1064, ORF1065, ORF1066, ORF1067, ORF1068, ORF1069, ORF1070, ORF1071, ORF1072, ORF1073, ORF1074, ORF1075, ORF1076, ORF1077, ORF1078, ORF1079, ORF1080, ORF1081, ORF1082, ORF1083, ORF1084, ORF1085, ORF1086, ORF1087, ORF1088, ORF1089, ORF1090, ORF1091, ORF1092, ORF1093, ORF1094, ORF1095, ORF1096, ORF1097, ORF1098, ORF1099, ORF1100, ORF1101, ORF1102, ORF1103, ORF1104, ORF1105, ORF1106, ORF1107, ORF1108, ORF1109, ORF1110, ORF1111, ORF1112, ORF1113, ORF1114, ORF1115, ORF1116, ORF1117, ORF1118, ORF1119, ORF1120, ORF1121, ORF1122, ORF1123, ORF1124, ORF1125, ORF1126, ORF1127, ORF1128, ORF1129, ORF1130, ORF1131, ORF1132, ORF1133, ORF1134, ORF1135, ORF1136, ORF1137, ORF1138, ORF1139, ORF1140, ORF1141, ORF1142, ORF1143, ORF1144, ORF1145, ORF1146, ORF1147, ORF1148, ORF1149, ORF1150, ORF1151, ORF1152, ORF1153, ORF1154, ORF1155, ORF1156, ORF1157, ORF1158, ORF1159, ORF1160, ORF1161, ORF1162, ORF1163, ORF1164, ORF1165, ORF1166, ORF1167, ORF1168, ORF1169, ORF1170, ORF1171, ORF1172, ORF1173, ORF1174, ORF1175, ORF1176, ORF1177, ORF1178, ORF1179, ORF1180, ORF1181, ORF1182, ORF1183, ORF1184, ORF1185, ORF1186, ORF1187, ORF1188, ORF1189, ORF1190, ORF1191, ORF1192, ORF1193, ORF1194, ORF1195, ORF1196, ORF1197, ORF1198, ORF1199, ORF1200, ORF1201, ORF1202, ORF1203, ORF1204, ORF1205, ORF1206, ORF1207, ORF1208, ORF1209, ORF1210, ORF1211, ORF1212, ORF1213, ORF1214, ORF1215, ORF1216, ORF1217, ORF1218, ORF1219, ORF1220, ORF1221, ORF1222, ORF1223, ORF1224, ORF1225, ORF1226, ORF1227, ORF1228, ORF1229, ORF1230, ORF1231, ORF1232, ORF1233, ORF1234, ORF1235, ORF1236, ORF1237, ORF1238, ORF1239, ORF1240, ORF1241, ORF1242, ORF1243, ORF1244, ORF1245, ORF1246, ORF1247, ORF1248, ORF1249, ORF1250, ORF1251, ORF1252, ORF1253, ORF1254, ORF1255, ORF1256, ORF1257, ORF1258, ORF1259, ORF1260, ORF1261, ORF1262, ORF1263, ORF1264, ORF1265, ORF1266, ORF1267, ORF1268, ORF1269, ORF1270, ORF1271, ORF1272, ORF1273, ORF1274, ORF1275, ORF1276, ORF1277, ORF1278, ORF1279, ORF1280, ORF1281, ORF1282, ORF1283, ORF1284, ORF1285, ORF1286, ORF1287, ORF1288, ORF1289, ORF1290, ORF1291, ORF1292, ORF1293, ORF1294, ORF1295, ORF1296, ORF1297, ORF1298, ORF1299, ORF1300, ORF1301, ORF1302, ORF1303, ORF1304, ORF1305, ORF1306, ORF1307, ORF1308, ORF1309, ORF1310, ORF1311, ORF1312, ORF1313, ORF1314, ORF1315, ORF1316, ORF1317, ORF1318, ORF1319, ORF1320, ORF1321, ORF1322, ORF1323, ORF1324, ORF1325, ORF1326, ORF1327, ORF1328, ORF1329, ORF1330, ORF1331, ORF1332, ORF1333, ORF1334, ORF1335, ORF1336, ORF1337, ORF1338, ORF1339, ORF1340, ORF1341, ORF1342, ORF1343, ORF1344, ORF1345, ORF1346, ORF1347, ORF1348, ORF1349, ORF1350, ORF1351, ORF1352, ORF1353, ORF1354, ORF1355, ORF1356, ORF1357, ORF1358, ORF1359, ORF1360, ORF1361, ORF1362, ORF1363, ORF1364, ORF1365, ORF1366, ORF1367, ORF1368, ORF1369, ORF1370, ORF1371, ORF1372, ORF1373, ORF1374, ORF1375, ORF1376, ORF1377, ORF1378, ORF1379, ORF1380, ORF1381, ORF1382, ORF1383, ORF1384, ORF1385, ORF1386, ORF1387, ORF1388, ORF1389, ORF1390, ORF1391, ORF1392, ORF1393, ORF1394, ORF1395, ORF1396, ORF1397, ORF1398, ORF1399, ORF1400, ORF1401, ORF1402, ORF1403, ORF1404, ORF1405, ORF1406, ORF1407, ORF1408, ORF1409, ORF1410, ORF1411, ORF1412, ORF1413, ORF1414, ORF1415, ORF1416, ORF1417, ORF1418, ORF1419, ORF1420, ORF1421, ORF1422, ORF1423, ORF1424, ORF1425, ORF1426, ORF1427, ORF1428, ORF1429, ORF1430, ORF1431, ORF1432, ORF1433, ORF1434, ORF1435, ORF1436, ORF1437, ORF1438, ORF1439, ORF1440, ORF1441, ORF1442, ORF1443, ORF1444, ORF1445, ORF1446, ORF1447, ORF1448, ORF1449, ORF1450, ORF1451, ORF1452, ORF1453, ORF1454, ORF1455, ORF1456, ORF1457, ORF1458, ORF1459, ORF1460, ORF1461, ORF1462, ORF1463, ORF1464, ORF1465, ORF1466, ORF1467, ORF1468, ORF1469, ORF1470, ORF1471, ORF1472, ORF1473, ORF1474, ORF1475, ORF1476, ORF1477, ORF1478, ORF1479, ORF1480, ORF1481, ORF1482, ORF1483, ORF1484, ORF1485, ORF1486, ORF1487, ORF1488, ORF1489, ORF1490, ORF1491, ORF1492, ORF1493, ORF1494, ORF1495, ORF1496, ORF1497, ORF1498, ORF1499, ORF1500, ORF1501, ORF1502, ORF1503, ORF1504, ORF1505, ORF1506, ORF1507, ORF1508, ORF1509, ORF1510, ORF1511, ORF1512, ORF1513, ORF1514, ORF1515, ORF1516, ORF1517, ORF1518, ORF1519, ORF1520, ORF1521, ORF1522, ORF1523, ORF1524, ORF1525, ORF1526, ORF1527, ORF1528, ORF1529, ORF1530, ORF1531, ORF1532, ORF1533, ORF1534, ORF1535, ORF1536, ORF1537, ORF1538, ORF1539, ORF1540, ORF1541, ORF1542, ORF1543, ORF1544, ORF1545, ORF1546, ORF1547, ORF1548, ORF1549, ORF1550, ORF1551, ORF1552, ORF1553, ORF1554, ORF1555, ORF1556, ORF1557, ORF1558, ORF1559, ORF1560, ORF1561, ORF1562, ORF1563, ORF1564, ORF1565, ORF1566, ORF1567, ORF1568, ORF1569, ORF1570, ORF1571, ORF1572, ORF1573, ORF1574, ORF1575, ORF1576, ORF1577, ORF1578, ORF1579, ORF1580, ORF1581, ORF1582, ORF1583, ORF1584, ORF1585, ORF1586, ORF1587, ORF1588, ORF1589, ORF1590, ORF1591, ORF1592, ORF1593, ORF1594, ORF1595, ORF1596, ORF1597, ORF1598, ORF1599, ORF1600, ORF

Db 141 SLILANKIKDKI-KEVAVNGEKLSDAVSGAPAAQWQPIILSNPKD--GVFAFK 197
QY 130 LVKE-----VGGTLOVS 141
Db 198 FIRDEKVIKAVVIGTNSVS 217
RESULT 13
ABBS8543
ID ABB68543 standard; Protein; 475 AA.
XX AC ABB68543;
XX 26-MAR-2002 (first entry)
XX Drosophila melanogaster polypeptide SEQ ID NO 32421.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABU12646.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX Disclosure; SEQ ID NO 32421; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL1840-ABL16175) and the encoded proteins
XX (ABBS7737-ABBS72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 475 AA;
Query Match 11.7%; Score 90; DB 22; Length 475;
Best Local Similarity 25.7%; Pred. No. 1.3;
Matches 37; Conservative 32; Mismatches 51; Indels 24; Gaps 7;
QY 7 DNVNLDEEQYELGFKEGGIGQTK--DOYLEGKGYGTFQFLIIGVIOELMKFWLSH 64
Db 47 DTAIRQAEKYLIDELQHONTRLTEQEAHAHELKLSAQ-----IQELK----- 93
QY 65 IDQYN-NSSLRNHLNLDIMAQISITNGKVEDYENIKAR-NKLRVIASITKETW 122
Db 94 -DQHYRNSLQHVNSLSIKTELNLTTGKR--QELERRLQIAQEKESLTSLEESD 150
QY 123 KIDSIDLNVKE----VGGTLOVSE 142

Db 151 RIHMLERHAREQETKLETTLOALE 174
RESULT 14
AAU12040
ID AAU12040 standard; Protein; 803 AA.
XX AC AAU12040;
XX 09-APR-2002 (first entry)
XX Clostridium difficile S-layer protein cell wall binding portion (ORF6).
XX Gene expression cassette; secretory leader sequence; anti-tumour therapy;
XX clostridial N-acetylmuramoyl-L-alanine amidase-like protein; cytostatic;
XX fibrinolysis; gangrenous tissue; necrotic tissue; infection;
XX S-layer protein cell wall binding portion; ORF6.
XX Clostridium difficile.
XX WO200194599-A1.
XX 13-DEC-2001.
XX 07-JUN-2001; 2001WO-SE01280.
XX 07-JUN-2000; 2000SE-0002139.
XX 26-APR-2001; 2001SE-0101479.
XX (SMIT-) SMITSKYDDSIINSTITUTET.
XX Burman LG, Akerlund T, Mukherjee K, Katagihallimath N;
XX WPI; 2002-122121/16.
XX New gene expression cassette comprising a secretory leader sequence
XX encoding a signal peptide from Clostridium difficile, useful e.g. for
XX presenting polypeptides on bacterial cell surface, or as anti-tumour
XX therapy -
XX Claim 6; Page 75-77; 113pp; English.
XX The present invention relates to a new gene expression cassette
XX comprising a secretory leader sequence encoding a signal peptide from
XX Clostridium difficile or signal peptides of analogous exported
XX clostridial N-acetylmuramoyl-L-alanine amidase-like proteins, linked
XX to a DNA sequence encoding a heterologous polypeptide. The gene
XX expression cassette is useful for presenting polypeptides on the surface
XX of bacterial cells, and/or secreting them into the surroundings of the
XX cells for mucosal immunisation, induction of immunological tolerance
XX and anti-tumour therapy in humans and animals, as vaccines, and for the
XX treatment of prevention of fibrinolysis in arterial or venous occlusion
XX and/or revitalising gangrenous or necrotic tissue in various diseases.
XX The gene expression cassette may also be used to produce in the gut,
XX e.g. peptides and enzymes for therapy and prophylaxis of various
XX diseases; single, fusion or multiple polypeptide antigens of microbial,
XX animal or mammalian origin for neonatal immune balancing, vaccination
XX against infections; and carrier molecules separated or fused to the
XX antigen to amplify or modulate immune response. Spores produced by
XX Clostridia, e.g. Clostridium difficile are useful for both industrial
XX production of vaccines and for local production of the desired
XX polypeptides at the body sites desired. AAU12037-AAU12048 represent
XX C. difficile S-layer protein cell wall binding portions encoded by
XX ORF1, ORF3, ORF5-7, ORF9, ORF11, ORF1, ORF4, ORF6, ORF8 or ORF1
XX respectively. The DNAs encoding these portions may be used in a
XX gene expression cassette of the present invention.
XX Sequence 803 AA;
Query Match 11.3%; Score 87.5; DB 23; Length 803;
Best Local Similarity 27.5%; Pred. No. 4.7;
Matches 36; Conservative 20; Mismatches 58; Indels 17; Gaps 6;

QY 23 EGIQGTGDOYLGKGYGTGQFRLI-IGYIQLMKFWLSHIDYNNSSSLRNHLNL 81
Db 570 ELRLGFKKYLIGGENSKVOTQLSNIGISVER-----ISGDRYKTSISLAQKLSI 625
QY 82 EDMAQISITNGDEVED-YEKIKKARNKLRVIAITKETWKIDSLDNLVKE----- 133
Db 626 KSV-SQVAVANGVGLADALSVGAAADNNPII--LTNEKSELQGADEFLNSSKITKSY 682
QY 134 -VGTLQVSEN 143
Db 683 IIGGTATLSSN 693

RESULT 15

AAV20137
ID AAY20137 standard; Protein; 239 AA.

AC AAY20137;

XX 19-JUL-1999 (first entry)

XX B. burgdorferi antigenic protein, t02A.aa BB002.

DE Antigenic protein; vaccine; Lyme disease; infection; detection.

XX Borrelia burgdorferi.

OS W09859071-A1.

XX 30-DEC-1998.

PD 18-JUN-1998; 98WO-US12718.

XX 03-SEP-1997; 97US-0057483.

PR 20-JUN-1997; 97US-0050359.

PR 22-JUL-1997; 97US-0053344.

PR 22-JUL-1997; 97US-0053377.

XX (HUMA-) HUMAN GENOME SCI INC.

PA (MEDI-) MEDIMUNE INC.

XX Choi GH, Erwin AL, Hanson MS, Lathigra R;

XX WPI; 1999-189980/16.

DR N-PSDB; AAX61834.

XX New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of diseases
PT caused by Borrelia, particularly Lyme disease
XX Claim 12; Page 211; 275pp; English.
PS This sequence represents a Borrelia burgdorferi (Bb) protein of the
CC invention, which is suitable for use in a vaccine. The Bb polypeptides
CC can be used in vaccines for eliciting protective antibodies to members of
CC the Borrelia genus, particularly for the use against Lyme disease in
CC humans and animals. They can be used for preventing or attenuating an
CC infection caused by a member of the Borrelia genus. The products can also
CC be used for detection of members of the Borrelia genus.

XX Sequence 239 AA;

Query Match 11.3%; Score 87; DB 20; Length 239;

Best Local Similarity 23.5%; Pred. NO. 1;

Matches 42; Conservative 32; Mismatches 47; Indels 58; Gaps 8;

QY 7 DNVNLESEQYELGFKGIGTQKQYLE--GKE-----YGVQ---TGFORFLIIGY 53

Db 31 ENIQNFKOKSGDLGASDEKFMGTATASELKAIGLEDRKNQYDIQIAKITNEENLIDTY 90

QY 54 IQE-----LMKFWLSHIDYNNSSSLRNHLNLEDIMAQISTNGDKVEDYEK 102

Db 91 ISAYELANENKWLKRFLLSLDY-----KKENIETLKEIL-----EKLNNYEN 136

QY 103 NIKKARNKLRVIA-----SITKETWKIDSLDNLVKEYGGTLQVSE 142
Db 137 DPKIAANFLYIALDIQLKLEKHLKSINEKLDITLSKENSRED-LEALLEQVKSALQLQE 194

Search completed: February 16, 2004, 14:09:50
Job time : 40 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 12:08:41 ; Search time 328 Seconds
(without alignments)
5020.073 Million cell updates/sec

Title: US-09-980-054a-11

Perfect score: 447

Sequence: 1 atgtcagatatagatataga.....accocgatgatgtggtga 447

Scoring table: IDENTITY_NUC

Gapop 10_0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	442.2	98.9	447	13	US-10-032-585-6062
C 2	70.4	15.7	4985	13	US-10-056-405-10
C 3	70.4	15.7	4985	15	US-10-094-240-10
C 4	65	14.5	65	13	US-10-032-585-2884
5	64.2	14.4	3673778	13	US-10-312-841-2
C 6	62.6	14.0	9539	13	US-10-240-453-53
C 7	62.6	14.0	9539	15	US-10-239-676-51
8	61.2	13.7	168575	13	US-10-178-194-1
C 9	60.2	13.4	17183	13	US-10-311-455-460
C 10	59.8	13.4	6109	13	US-10-311-455-299
C 11	59.8	13.4	12237	13	US-10-311-455-2331
C 12	59.4	13.3	446	10	US-09-960-352-3400
C 13	59.2	13.2	12142	13	US-10-311-455-1646
C 14	58.2	13.0	3673778	13	US-10-312-841-2
C 15	57.8	12.9	516	10	US-09-960-352-5785

ALIGNMENTS

RESULT 1
US-10-032-585-6062
; Sequence 6062, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6062
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-6062

Query Match	98.9%	Score 442.2;	DB 13;	Length 447;
Best Local Similarity	99.3%	Pred. No. 4.1e-67;		
Matches	444;	Conservative	0;	Mismatches 3;
Indels	0;	Caps	0;	
Qy	1	ATGTCAGATATAGATATAGATATATGTTAAATTTAGAGAGAACAAATATGTAATAGGA	60	
Db	1	ATGTCAGATATAGATATAGATATGTTAAATTTAGAGAGAACAAATATGTAATAGGA	60	
Qy	61	TTTAAAGAGGTCAATACAGGACAAAGATCAATATTTAGAGAGAAAGATATGTT	120	
Db	61	TTTAAAGAGGTCAATACAGGACAAAGATCAATATTTAGAGAGAAAGATATGTT	120	
Qy	121	TATCAAACTGGATTTCACGATTTTAAATCATTGGTTATTTATTCAGAAATTAATGAAATTT	180	
Db	121	TATCAAACTGGATTTCACGATTTTAAATCATTGGTTATTTATTCAGAAATTAATGAAATTT	180	
Qy	181	TGTTATCCCATATAGATCAATATATACTCTTCTCTCTCGGATCATTTGATATAT	240	

Db 181 TGGTTATCCCATATAGATCAATATAAATCTCTTCCTTCGGAATCATTTGAATAAT 240
Qy 241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAGTTTGAAGATTAT 300
Db 241 TTGGAAGATATTATGGCACAAATTTCTATAACGAATGGAGATAAAGAGTTTGAAGATTAT 300
Qy 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAAGAGTATAGCTAGTATACTAAAGAA 360
Db 301 GAAAAAATATTAAAAAGGCAAGAAATAAATTAAAGAGTATAGCTAGTATACTAAAGAA 360
Qy 361 ACTTGGAAATTTGATTCATTTGGTAATTTTGGTGAAGAAAGTATAGTGAATTTCAAGTT 420
Db 361 ACTTGGAAATTTGATTCATTTGGTAATTTTGGTGAAGAAAGTATAGTGAATTTCAAGTT 420
Qy 421 AGTGAACACCCGATGATATGTGTGA 447
Db 421 AGTGAACACCCGATGATATGTGTGA 447

RESULT 2
US-10-056-405-10/c
; Sequence 10, Application US/10056405
; Publication No. US20030166013A1
; GENERAL INFORMATION:
; APPLICANT: ZWIBEL, LAURENCE J.
; TITLE OF INVENTION: MOSQUITO OLFATORY GENES, POLYPEPTIDES, AND METHODS OF
; FILE REFERENCE: N7841
; CURRENT APPLICATION NUMBER: US/10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-056-405-10

Query Match 15.7%; Score 70.4; DB 13; Length 4985;
Best Local Similarity 51.0%; Pred. No. 0.0073;
Matches 196; Conservative 0; Mismatches 181; Indels 7; Gaps 1;
Qy 4 TCAGATATAGATATAGATTAATGTTTAAATTTAGAGAGCAAGCAATATGAATGAGATT 63
Db 2850 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATAATA 2791
Qy 64 AAAGAAGTCAATATCAAGGCAACAAAGATCAATATTTAGAGAGCAAGCAATATGAAT 123
Db 2790 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2731
Qy 124 CAAACTGGATTTCAACGATTTTAAATCAATGTTTATTTCAAGAAATTAATGAAATTTGG 183
Db 2730 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2678
Qy 184 TTATCCCATATAGATCAATATAAATCAATCTTCTTCACCTTCGGAATCATTTGAATAATTG 243
Db 2677 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2618
Qy 244 GAAGATATTATGGCAAAATTTCTATAACGAATGGAGATAAAGAGTTTGAAGATTATGAA 303
Db 2617 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2558
Qy 124 CAAACTGGATTTCAACGATTTTAAATCAATGTTTATTTCAAGAAATTAATGAAATTTGG 183
Db 2730 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2678
Qy 184 TTATCCCATATAGATCAATATAAATCAATCTTCTTCACCTTCGGAATCATTTGAATAATTG 243
Db 2677 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2618
Qy 244 GAAGATATTATGGCAAAATTTCTATAACGAATGGAGATAAAGAGTTTGAAGATTATGAA 303
Db 2617 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2558
Qy 304 AAAAATATTAAAAAGGCAAGAAATAAATTAAAGAGTATAGCTAGTATACTAAAGAACT 363
Db 2557 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2498
Qy 364 TGGAAATTTGATTCATTCGATAAT 387
Db 2497 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2474

RESULT 3
US-10-094-240-10/c
; Sequence 10, Application US/10094240
; Publication No. US20030082637A1
; GENERAL INFORMATION:
; APPLICANT: ZWIBEL, LAURENCE J.
; TITLE OF INVENTION: ARRESTIN GENE, POLYPEPTIDE, AND METHODS OF USE THEREOF
; FILE REFERENCE: N8289
; CURRENT APPLICATION NUMBER: US/10/094,240
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 10/056,405
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/264,649
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 4985
; TYPE: DNA
; ORGANISM: Anopheles gambiae
US-10-094-240-10

Query Match 15.7%; Score 70.4; DB 15; Length 4985;
Best Local Similarity 51.0%; Pred. No. 0.0073;
Matches 196; Conservative 0; Mismatches 181; Indels 7; Gaps 1;
Qy 4 TCAGATATAGATATAGATTAATGTTTAAATTTAGAGAGCAAGCAATATGAATGAGATT 63
Db 2850 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2791
Qy 64 AAAGAAGTCAATATCAAGGCAACAAAGATCAATATTTAGAGAGCAAGCAATATGTTTAT 123
Db 2790 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2731
Qy 124 CAAACTGGATTTCAACGATTTTAAATCAATGTTTATTTCAAGAAATTAATGAAATTTGG 183
Db 2730 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2678
Qy 184 TTATCCCATATAGATCAATATAAATCAATCTTCTTCACCTTCGGAATCATTTGAATAATTG 243
Db 2677 TAAATATTAATAATATTAATAATAATAATAATAATAATAATAATAATAATAATA 2618
Qy 244 GAAGATATTATGGCAAAATTTCTATAACGAATGGAGATAAAGAGTTTGAAGATTATGAA 303
Db 2617 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2558
Qy 304 AAAAATATTAAAAAGGCAAGAAATAAATTAAAGAGTATAGCTAGTATACTAAAGAACT 363
Db 2557 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2498
Qy 364 TGGAAATTTGATTCATTCGATAAT 387
Db 2497 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2474

RESULT 4
US-10-032-585-2884/c
; Sequence 2884, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2884
; LENGTH: 65

```
; TYPE: DNA
; ORGANISM: Candida albicans
US-10-032-585-2884

Query Match      14.5%; Score 65; DB 13; Length 65;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCAGATATAGATATAGATATATTAATTTAGAGAGAACATATGAATTAGGATTT 63
Db 65 TCAGATATAGATATAGATATATTAATTTAGAGAGAACATATGAATTAGGATTT 6
QY 64 AAAGA 68
Db 5 AAAGA 1

RESULT 5
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match      14.4%; Score 64.2; DB 13; Length 3673778;
Best Local Similarity 54.2%; Pred. No. 0.57;
Matches 173; Conservative 0; Mismatches 143; Indels 3; Gaps 2;

QY 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAACATATGAATTAGGA 60
Db 554727 ATATTAAATTTAGAAATTTTAATTAATATATTTAAAGGAAAAAGTTAATTTAA 554786
QY 61 TTTAAAGAGGTCAAATACAGGAACAAAGATCAATATTTAGAAG-GAAAAAGATATGG 119
Db 554787 ATTTATTTATATATATATATTTATATAGGATTTATTTAGATATTTAAGATAGTT 554846
QY 120 TTATCAACTCGATTTCACGATTTTAAATCATCTGGTTATATTCAGAAATTAATGAATTT 179
Db 554847 TTATAGGAAATGATTAAAGATGTGAAGGGTAGTTTATATAAGGGATATATAAAGA 554906
QY 180 TTGGTTATCCCATATAGATCAATATAAATCACTCTTCTTCACTTCGGAATCATTTGAATAA 239
Db 554907 TTAATTAACGTATAAAGTTAAGTTTATATATAGTTTATAGAGAATTTGAAATTAGATAG 554966
QY 240 TTGGAGATATATGCGCAAAATTTCTATACGAATGGAGATAAAGAGTTGAAGATTA 299
Db 554967 TTTTGAATAATAT- -TTTAAATTTAATATAGATAAATTTGAAAAAGTTTATATTA 555024
QY 300 TCAAAAAATATATAAAG 318
Db 555025 TTAATAAATTAATGAGGATG 555043

RESULT 6
US-10-240-453-53/c
; Sequence 53, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
```

```
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
; TITLE OF INVENTION: With DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240,453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 350
; SEQ ID NO 53
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-453-53

Query Match      14.0%; Score 62.6; DB 13; Length 9539;
Best Local Similarity 47.8%; Pred. No. 0.19;
Matches 182; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 4 TCAGATATAGATATAGATATATTAATTTAGAGAGAACATATGAATTAGGATTT 63
Db 6192 TCAAAAAATTTAAATATATACCATTAATTTTAAAAAACAATATATAACGAATTTAT 6133
QY 64 AAAGAAGGTCAAATACAGGAACAAAGATCAATATTTAGAACGAAAGAAATATGGTTAT 123
Db 6132 AAAACAAAAACGAACTTAAAAATTTCAAAAAACACATTTTAACTCAAAAAATATAAA 6073
QY 124 CAAACTGGATTTCAACGATTTTAAATCAATTTGTTTATTTCAAGAATTAATGAATTTTG 183
Db 6072 AACAAACGAATTTAAAAATCGAAACGAATTTAAAAATCAAAAAATTAATAAATAAATA 6013
QY 184 TTATCCCATATAGATCAATATAATAACTCTTCTTCTCGGAATCATTTGAATAATTTG 243
Db 6012 ATATAAAATCAAATCTTAAATATCAAAAAATCTAAAAAATTTCTTCTAA 5953
QY 244 GAAGATATTTATGCAACAAATTTCTATAACGAATGAGATAAAGAGTTGAGATTATGAA 303
Db 5952 AAAAAAATAAAAATTTAAAAATCAAAATCTATTTAAACAAACAAATATAACTAATAAA 5893
QY 304 AAAAAATTTAAAAAGGCAAGAAATAAATTAAGAGTGTAGTAGTAACTAAAGAAACT 363
Db 5892 AAAAAATTTAAAAACCAAAATTAATTAATAAABAAAAAATTTAAACCTTAAACAAAAAAT 5833
QY 364 TGGAAATTCATTCATTTGGAT 384
Db 5832 CAAAAAACAATAAATAATATAT 5812

RESULT 7
US-10-239-676-51/c
; Sequence 51, Application US/10239676
; Publication No. US20030082609A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239, 676
```

; CURRENT FILING DATE: 2002-09-24
; PRIOR APPLICATION NUMBER: PCT/EP01/03968
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-04-06
; 2000-04-06
; 2000-04-07
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 228
; SEQ ID NO 51
; LENGTH: 9539
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
US-10-239-676-51

Query Match 14.0%; Score 62.6; DB 15; Length 9539;
Best Local Similarity 47.8%; Pred. No. 0.19; Indels 0; Gaps 0;
Matches 182; Conservative 0; Mismatches 199; Indels 0; Gaps 0;
QY 4 TCAGATATAGATAPAGATATGTTTAAATTTAGAGAGACAAATATGATTTAGGATTT 63
DB 6192 TCAAAAAAATTTAAATATTAACCAATAAAATTTATAAAAAACAATATTAACGAAATTTAT 6133
QY 64 AAGAAGGTCAATACAGGACAAAGATCAATATTTAGAGGAAAGATATGTTAT 123
DB 6132 AAAACAAAAGAACTTAAAAATTCAAAAACACATTTTAACTCAAAAATTAATATAA 6073
QY 124 CAAACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAGAAATTTAATGAAATTTTGG 183
DB 6072 AACAAAGAAATTTAAAAATTCGAAACGAAATTTAAAAATCAAAAATTAATAAATA 6013
QY 184 TTATCCCATATAGATCAATATAATACTTCTTCACTTCGGATCATTTGATTAATTTG 243
DB 6012 ATATAAAAATCAAAATTCATAAAATCATAAAAATCTAAAAAATTCAAAAATTTTCTAA 5953
QY 244 GAAGATATTATGGCACAAATTTCTATACGAATGGAGATTAAGAAAGTTGAAGATTTATGAA 303
DB 5952 AAAAAAATAAAATTTAAATCAAAATCTATTTAAACAAACAAATATACTAATAA 5893
QY 304 AAAAATATTAAGGACGAGAAATTAATTAAGAGTGATAGTAGTATTAAGTAAAGAACT 363
DB 5892 AAAAAATTTAAAAACCAAAATTAATTAATTAATAAAAAAACTTAAAACTAAAAAATTT 5833
QY 364 TGGAAATTTGATTCATTTGGAT 384
DB 5832 CAAAAACAATAAATATAT 5812

RESULT 8
US-10-178-194-1
; Sequence 1, Application US/10178194
; Publication No. US20030166904A1
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Gulcher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/10/178,194
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US/09/426,290
; PRIOR FILING DATE: 1999-10-25
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:

; NAME/KEY: CDS
; LOCATION: (21181)...(21403)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (95252)...(95430)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101753)...(101996)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (110324)...(110439)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (124058)...(124278)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (127009)...(127130)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128910)...(129139)
US-10-178-194-1
Query Match 13.7%; Score 61.2; DB 13; Length 168575;
Best Local Similarity 48.1%; Pred. No. 0.77; Indels 1; Gaps 1;
Matches 203; Conservative 0; Mismatches 218; Indels 1; Gaps 1;
QY 8 ATATAGATATAGATATGTTTAAATTTAGAGAGAAACAATAT-GAATTAGGATTTAAA 66
DB 108964 ATATATAATATATAATGTTTATAGTTATATATAATATATATAATATAATATGTA 109023
QY 67 GAAGTCAAAATACAGGAAACAAGATCAATATTTAGAGGAAAGAAATATGGTTACAA 126
DB 109024 TTATATGTTATATAATATTTACATATAAATATCTATTATATATATATATATATATA 109083
QY 127 ACTGCAATTTCAACGATTTTAAATCAATGTTTATATTCAGAAATTTAATGAAATTTTGGTTA 186
DB 109084 TCTAATATATAATATAATACTTATTATATATAATATAATACTATATATATATA 109143
QY 187 TCCCATATAGATCAATATAAATCACTTCTTCTCCTCGAATCACTTTGAATAATTTGGAA 246
DB 109144 TAATATATAATATAATATAAATCACTTATTATATATATATATATATATAATAAATAT 109203
QY 247 GATATTATGGCACAAATTTCTATACGAATCGAGATTAAGAAAGTTGAAGATTTATGAAAA 306
DB 109204 GTACTATATTAATATATGATATATCTAATTAATTAATATACATATATAAATTTATAATA 109263
QY 307 AATATTAAGGACGAGAAATTAATTAAGAGTGATAGCTAGTATAACTTAAGAAACTTGG 366
DB 109264 TATAATGATTATATATATATAATAATAATAATAATAATAATAATAATACTGAAATTTA 109323
QY 367 AAAATTTGATTCATTTGGATAATTTGGTGAAGAAAGTAGTGAACTTTTACAAGTTAGTCAA 426
DB 109324 TTTATGTTTATGATACATATAAAGTAGGAGGAAATATATATATATATATATATAGTAGGAGAA 109383
QY 427 AA 428
DB 109384 TA 109385

RESULT 9
US-10-311-455-460/c
; Sequence 460, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK Alexander
; APPLICANT: PIEFENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537

;; PRIOR FILING DATE: 2001-07-02
;; PRIOR APPLICATION NUMBER: DE 10032529.7
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: DE 10043826.1
;; PRIOR FILING DATE: 2000-09-01
;; NUMBER OF SEQ ID NOS: 2424
;; SEQ ID NO 460
;; LENGTH: 17183
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-460

Query Match 13.5%; Score 60.2; DB 13; Length 17183;
Best Local Similarity 49.9%; Pred. No. 0.59;
Matches 185; Conservative 0; Mismatches 178; Indels 8; Gaps 1;

QY 8 ATATGATATAGTATGTTAAATTTAGAGAGAACAAATATGAATAGGATTTAAAG 67
DB 4235 ATAAAAAATAATCATTAATACTACTCTAAAAAAATTTAAAAAACTATTAACCCCTCTAAAC 4176
QY 68 AAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTTATCAAA 127
DB 4175 AAAAAAACAACATTTACAAAACATATATACAAAACATATATCTTAATAAAAT 4116
QY 128 CTGGATTTCAAGATTTTAAATCATTTGTTATTAATCAAGAAATTAATGAAATTTGGTTAT 187
DB 4115 CAATATTATTAATATAAATACTACATAATAAAAAATAATTAATTAATAAATAAATAA 4056
QY 188 CCATATAGATCAATATAATACTCTCTTCCTCACTTCGGAATCATTTGAATATTTGGAG 247
DB 4055 CTACTTATAAATAATATATACATTTTTCCTATAAACAATAAATAATCAATAAAT 3996
QY 248 ATATTANGCAAAATTTCTATAAC-----GAATGAGATAAAGAGTTGAAGATTA 299
DB 3995 TTTTAATATACAAATAACATAACACATTTATATTTTAAAAAATAAATATACATAA 3936
QY 300 TGAATAAATATTAAGAGCGAAGAAATAAATTAAGATGATAGTAGTATTAACATAAAGA 359
DB 3935 TATTAATAATATACCAACCCAAAAAACAATTAATAAATAAATATATATCATTTCAATA 3876
QY 360 AACTTGGAAAA 370
DB 3875 ATCCAATAAAA 3865

RESULT 10
US-10-311-455-299/c
; Sequence 299, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 299
; LENGTH: 6109
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

;; NAME/KEY: unsure
;; LOCATION: 214, 2796..2797, 4347
;; OTHER INFORMATION: n is a or g or c or t
US-10-311-455-299

Query Match 13.4%; Score 59.8; DB 13; Length 6109;
Best Local Similarity 46.3%; Pred. No. 0.52; 227; Indels 0; Gaps 0;
Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 6 AGATATAGATATAGTATGTTAAATTTAGAGAGAACAAATATGAATAGGATTTAA 65
DB 2081 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2022
QY 66 AGRAGGTCAAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTTATCA 125
DB 2021 ATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1962
QY 126 AACTGGATTTCAAGATTTTAAATCATTTGTTATATTTCAAGAAATTAATGAAATTTTGGTT 185
DB 1961 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1902
QY 186 ATCCATATAGATCAATATATTAATCACTCTTCCTCACTTCGGAATCATTTGAATTTTGGGA 245
DB 1901 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1842
QY 246 AGATATTATGCGCAAAATTTCTATAACGAATCGAGATATAAAGAGTTTGAAGATTATGAAAA 305
DB 1841 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1782
QY 306 AATATTAAAAAGCGCAAGAAATAAATAAGAGTGATAGCTAGTATTAAGAGAACTTG 365
DB 1781 ATAAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1722
QY 366 GAAATTTGATTCATTTGGATAATTTGGTGAAGAGTAGGTGGAACTTTTACAAGTTAGTGA 425
DB 1721 ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1662
QY 426 AAA 428
DB 1661 ATA 1659

RESULT 11
US-10-311-455-2331/c
; Sequence 2331, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determ
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2331
; LENGTH: 12237
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2331

Query Match 13.4%; Score 59.8; DB 13; Length 12237;
Best Local Similarity 46.3%; Pred. No. 0.63;


```
/
/
/ LENGTH: 615 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: PAG1074RP
/ US-08-998-416-186

Query Match
Best Local Similarity 12.5%; Score 55.8; DB 3; Length 615;
Matches 174; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

QY 8 ATATAGATATAGATAATGATTAATTAATTTAGAGAGACAAATATGATTAAGTTTAAAG 67
Db 603 ATATTTTATTAATTAATTAATTAATGATTAATTAATTAATTAATTAATTAATTA 544
QY 68 AAGGTCAAATACAGGACAAAGATCAATATTTAGAGAGAGAGAGAGATGATGTTACAAA 127
Db 543 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAGAAA 484
QY 128 CTGGATTTCACAGATTTTAAATCAATGTTTATTAATCAAGAAATTAATGAAATTTTGGTTAT 187
Db 483 TTAAAGTTTAAATTAATTTTAAATTAATTTCTTATTAAGAGTTTAAATTAATTAATCAAC 424
QY 188 CCCATATAGATCAATATAATAACTCTTCTCACTTCGGAAATCATTTGAAATATTTTGGAG 247
Db 423 ATAATATTTTAAATTAATAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 364
QY 248 ATATTATGACACAAATTTCTATAACGATGAGAGATAAGAGTTTCAAGATTAATGAGAAA 307
Db 363 TAAATCTTTATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 304
QY 308 ATATTAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATTAATACTAAAGAAATCTTGA 367
Db 303 TTATTGATAATCTATTTAATTAATTTTATTAAGAAATAATTAATTAATTAATTAATTTTAA 244
QY 368 AAATTCATCA 378
Db 243 TAACTAAATTA 233

RESULT 4
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5933827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
/
/
/ NAME: Israelsen, Ned
/ REGISTRATION NUMBER: 29,655
/ REFERENCE/DOCKET NUMBER: NIH121.001CP1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 235-8550
/ TELEFAX: (619) 235-0176
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 19124 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ US-08-487-826B-13

Query Match
Best Local Similarity 12.4%; Score 55.6; DB 2; Length 19124;
Matches 185; Conservative 0; Mismatches 174; Indels 5; Gaps 2;

QY 9 TATAGATATAGATATGATTAATTTAGAGAGAGACAAATATGATTAAGTTTAAAGA 68
Db 15636 TATAAAAAAATAAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15695
QY 69 AGGTCAATACAGGACAAAGATCAATATTTAGAGAGAGAGAGAGAGAGATAT--GGTTATCA 125
Db 15696 TCATAAAAACAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 15755
QY 126 AACTGGATTTCAAGGATTTTATCATCTGTTTATATTTCAAGAGATTAATGAAATTTTGGTT 185
Db 15756 AAAAAAATTTAAAAAATGTTAAAAAATAATATATACATATAAAAAAATAAAAAATTTATTT 15815
QY 186 ATCCATATAGATCAATATAATAACTCTTCTTCACTTCGGAAATCATTTGAAATTAATTGGA 245
Db 15816 AATATAAAAAAATAATATAATAAAAAATTTATTAATAAAAAAATAATTAATAAAAA 15875
QY 246 AGATATTATGCGCAAAATTTCTATAACGATGAGAGATAAGAGTTTGAAGATTTATGAAA 305
Db 15876 AAATTTAATTAATAAAAAAATAAAAAAATTTAAAAA--ATTTAATGAATAAAAAAATAAAA 15933
QY 306 AAATATTAAAAAGGCAAGAAATAAATTAAGAGTGATAGCTAGTATTAACATAAGAGAACTTG 365
Db 15934 AATTTAATTAATAAAAAAATAAAAAAATAAATAAATAATTAATACATGACATATACATATA 15993
QY 366 GAAA 369
Db 15994 TATA 15997

RESULT 5
US-09-801-861-3/c
; Sequence 3, Application US/09801861
; Patent No. 6492154
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001098
; CURRENT APPLICATION NUMBER: US/09/801,861
; CURRENT FILING DATE: 2001-03-09
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53332
; TYPE: DNA
; ORGANISM: Human
; US-09-801-861-3

Query Match
Best Local Similarity 12.2%; Score 54.6; DB 4; Length 53332;
Matches 178; Conservative 0; Mismatches 174; Indels 9; Gaps 1;
```


QY 126 AACTGGATTTCAACGATTTTAAATCATGTTTATATTCACGAATTAATGAATTTGGTT 185
 Db 1893 AAATGTGGAAGATATTTATGAAAATAATGAATACAGATTCGATTTATTAATAAGAAAT 1834
 QY 186 ATCCCATATGATCAATATATTAATCACTCTTCTCACTTCGGAATCATTTGGAATTTGGA 245
 Db 1833 ATATGATAAGTCAAAAAATAATATATAATCTGTTTTATGAATTCGTAAAAAGATAGA 1774
 QY 246 AGATATTTATGCGACAAATTTCTATAACGAATGCGAGATAGAGAGTTGGAAGATTAAGAAA 305
 Db 1773 TGATACTATACACATTTTCTGATCGATCAATATTAATGATGAAGGAATATAATAATAT 1714
 QY 306 AATATTTAAAAAGCAAGAAATAAATTTAA 334
 Db 1713 TAATAATGATATTAAGGAAGGATAACAA 1685
 RESULT 8
 US-08-998-416-288/c
 ; Sequence 288, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgen
 ; APPLICANT: Knechtle, Philipp
 ; APPLICANT: Reibschung, Corinne
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
 ; NUMBER OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: No. 6239264th Carolina
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/998,416
 ; FILING DATE: 24-DEC-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 0016/97
 ; FILING DATE: 31-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 288:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 837 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ORIGINAL SOURCE:
 ; ORGANISM: PAG1241RP
 ; US-08-998-416-288

Query Match 11.7%; Score 52.4; DB 3; Length 837;
 Best Local Similarity 48.1%; Pred. No. 0.015;
 Matches 178; Conservative 0; Mismatches 191; Indels 1; Gaps 1;

QY 9 TATAGATATAGATATGTTTAAATTTAGAGAGAACAAATATGATTTAGGATTTAAGA 68
 Db 601 TTTATTAATTAATTAATTAATGATATAATATTAATTTAAATTAATTTAATTAATAA 542
 QY 69 AGGTCAATACAGAGAACAAAGAGATCAATATTTAGAGAGAGAGAGAGATTTATCAAC 128
 Db 541 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 483
 QY 129 TGGATTTCAACGATTTTAAATCACTTTGTTTATATTCAGAGATTAATGAAATTTTGGTTATC 188
 Db 482 AAAGTTAAATTAATTTTAAATTAATTAATTTTATATAAAGATTAATTAATTAATTAATTAAT 423
 QY 189 CCATATAGATCAATATAATACTCTTCTTCATCTTCGGAATCATTTGGAATTTTGAAGA 248
 Db 422 TAATATTTATAAATAAGATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 363
 QY 249 TATTATGGCACAATTTCTATAACGAATGAGAGATAAGAGAGATTTGAAGATTAAGAAAAA 308
 Db 362 AAATCTTTTATAATAAATAATATATTTTAAATAAACAATTAATAATAATAATAATAATAA 303
 QY 309 TATTAAAAAGGCAAGAAATAAATAAGAGATGATAGTAGTATACTAAAGAACTTGAA 368
 Db 302 TATTGATATCTTATTATAATAATTTTAAAGAAAAATAATAATAATAATAATAATAATAA 243
 QY 369 AATTGATTTCA 378
 Db 242 AACTAATTTA 233
 RESULT 9
 US-08-998-416-1137/c
 ; Sequence 1137, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgen
 ; APPLICANT: Knechtle, Philipp
 ; APPLICANT: Reibschung, Corinne
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
 ; NUMBER OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: No. 6239264th Carolina
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/998,416
 ; FILING DATE: 24-DEC-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 0016/97
 ; FILING DATE: 31-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 1137:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 636 base pairs


```
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (600592)..(600592)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
```

```
; NAME/KEY: misc feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
; US-08-916-421B-1
```

Query Match 11.0%; Score 49; DB 4; Length 1664976;
Best Local Similarity 45.5%; Pred. No. 0.15;
Matches 175; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

```
Qy 36 AGAAGAAGAACCAATATGAATTTAGGATTTAAAGAAGTCAATACCAAGGAACCAAGATCA 95
Db 844063 AGAAAAAGAAAAACCTAAAGAGATTAAGAAAAAGAACTGAAGAGAAAT 844004

Qy 96 ATATTTAGAGAAAGAAATATGGTTATCAAACTGGATTTCAACGATTTTAAATCATTTGG 155
Db 844003 AGAAAAAGAAAAATAGAAATTTGTTAAGAAAGAAAGAAACAAATTTATAAAAAATC 843944

Qy 156 TTATATTCAGAAATTAATCAAAATTTGGTTATCCCATATAGATCAATATAAATCTCTTC 215
Db 843943 TGAATGAAGATGTTGAAGAGAAATTAACAACAACTAATTTCCAAAGAGAGAAAAAGAGA 843984

Qy 216 TTCACCTCCGAATCATTTGAATTAATTTGGAAGATATTATGGCAAAATTTCTTATAACGAA 275
```

Db 843883 TTTCATGCTGAGAGGCTAAAGATATGACCATACAAAATAAAGAGAGTGTAA 843824
Qy 276 TGGAGATAAAGACTTGAAGATTATGAAAAAATATTAAGAGCAGAAATAAATTAAG 335
Db 843823 TAGTAGATAAATGGATAGCTAAGACATCGATCGCTGATGAGATATATGAGATTC 843764
Qy 336 AGTGATAGCTAGTATATACTAAAGAACTTGGAAATTTGATTCATTTGATATTTGGTGAA 395
Db 843763 AGACGTGTCTGGAAATCCACATCGACTGGAACCTATTGAGGACTTCGTTAAATCTTAG 843704
Qy 396 AGAAGTAGTGGAACTTTACAGTT 420
Db 843703 AGACAGATTGAAAGATTAAAGTT 843679

RESULT 12

US-09-601-198-70
; Sequence 70, Application US/09601198
; Patent No. 6531583
; GENERAL INFORMATION:
; APPLICANT: Cassell, Gail H.
; APPLICANT: Chen, Ellison Y.
; APPLICANT: Glass, Jennifer S.
; APPLICANT: Glass, John I.
; APPLICANT: Heiner, Cheryl R.
; APPLICANT: Lefkowitz, Elliot
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
; TITLE OF INVENTION: UREA PLASMA
; FILE REFERENCE: UAB-13452/22
; CURRENT APPLICATION NUMBER: US/09/601.198
; CURRENT FILING DATE: 2000-12-08
; PRIOR APPLICATION NUMBER: 60/073,189
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 181
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO: 70
; LENGTH: 2427
; TYPE: DNA
; ORGANISM: Ureaplasma urealyticum
US-09-601-198-70

Query Match 10.8%; Score 48.2; DB 4; Length 2427;
Best Local Similarity 47.9%; Pred. No. 0.11; 183; Indels 3; Gaps 1;
Matches 171; Conservative 0; Mismatches 183; Indels 3; Gaps 1;
Qy 72 TCAATACAGGACAAAGATCAATATTTAGAGGAAAAGAAATATGTTTCAAACTGG 131
Db 1701 TGAAGAATCCAGATCTGATAAGATTACTAGTTGAAGAGCGTGAATATCTCTTG 1760
Qy 132 ATTTCAACGATTTTATCATGTTATATTTCAAGAAATTAATGAATTTGGTTATCCCA 191
Db 1761 AATTAAATACATCACTAATTAATATATATATCAACTTCCCAATTTTAAATATCTTATTTAA 1820
Qy 192 TATAGATCAATATAATACTCTTCTCACTTCGGAATCATTTGAATTAATTTGGAAGATAT 251
Db 1821 TCAATCTGAATGATTTATCAATCAATATCTAAGAAATGATGATTAATAAATA 1880
Qy 252 TATGCACAAAATTTCTATAAGAAATGAGATTAAGAGTTGAAGATTTGAAGAAATAT 311
Db 1881 TTTACTACAAATTTTGAACCATATCAAGATAAAGAAATGGAATTTTAAATTTAAACA- 1937
Qy 312 TAAAGGCAAGAAATAATTAAGAGTGATAGTAGTATTAATAAGAAATCTTGGAAAT 371
Db 1938 TAAAGATTGACAGATGCGGACAGATGTTGCTGATTTATTAATAATGACAGATGATATAG 1997
Qy 372 TGATTCATTTGGATATTTTGGTGAAGAGATAGGTGGAACTTTTACAAGTTAGTGA 428
Db 1998 CGACCAAAAAGAAATATTATTTAACGGCTATGTTTAAATAATCAACGAGCTTATATTA 2054

RESULT 13

US-08-446-855A-1/c

; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S
; APPLICANT: Flores, Maria V
; APPLICANT: O'Sullivan, William J
; TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
; TITLE OF INVENTION: Phosphate synthetase II
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye PC
; STREET: 1100 No. 5849573th Glebe Road, 8th Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA: US/08/446.855A
; APPLICATION NUMBER: US/08/446.855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic
US-08-446-855A-1

Query Match 10.7%; Score 48; DB 2; Length 8920;
Best Local Similarity 46.0%; Pred. No. 0.14;
Matches 198; Conservative 0; Mismatches 230; Indels 2; Gaps 1;
Qy 8 ATATAGATATAGATTAATGTTAAATTTAGAGAGAAACAATATGAATAGGATTTAAAG 67
Db 877 AGATATATATATCTTTTAATAACAACCTTTCTGATGTTAAAGAAATAAACTGTTAAGACC 818
Qy 68 AAGGTCAAATACAAGAACAAAGATCAATATTTAGAGAGAAAGAAATATGTTTATCAA 127
Db 817 TATGATTCAGAAATATCCCAATTAATATATATATATATATATATATATATATATATAT 758
Qy 128 CTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAGAAATTAATGAATTTTGGTTAT 187
Db 757 ATTATTTTTCCTCAATTTCTTTTCTTTTATATACATTTAT-ATTTATGTTTAAATA 700
Qy 188 CCCATATAGATCAATATATATCTTCTTCACTTCGGAATCATTTGATTAATTTTGGAG 247
Db 699 TTTATAAATTTACATATACAGTTTCATTTTTCATATGTAATTTTCTTTTCTTTTCTTTT 640
Qy 248 ATATTATGTCACAAATTTCTATAAGCAATGGAGATAAGAACTTGAAGATTAATGAATA 307
Db 639 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 580
Qy 308 ATATTAAAGCAAGAAATAAATTAAGAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAG 367
Db 579 AAATAAATAAATGAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 520
Qy 368 AAATTCATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 427
Db 519 ATTGGGAGTATATATATTTGTTATATATATATATATATATATATATATATATATATATAT 460

QY 428 ACCCGATGA 437
 Db 459 AACTATATTA 450

RESULT 14

US-09-150-741-1/c
 ; Sequence 1, Application US/09150741
 ; Patent No. 6183996
 ; GENERAL INFORMATION:
 ; APPLICANT: Stewart et al.
 ; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
 ; Patent No. 6183996
 ; FILE OF INVENTION: Synthetase II
 ; CURRENT APPLICATION NUMBER: US/09/150,741
 ; CURRENT FILING DATE: 1998-09-10
 ; EARLIER APPLICATION NUMBER: PL6380
 ; EARLIER FILING DATE: 1992-12-16
 ; EARLIER APPLICATION NUMBER: AU93/00617
 ; EARLIER FILING DATE: 1993-12-02
 ; EARLIER APPLICATION NUMBER: 08/446,855
 ; EARLIER FILING DATE: 1995-07-06
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 8920
 ; TYPE: DNA
 ; ORGANISM: Plasmodium falciparum
 US-09-150-741-1

Query Match 10.7%; Score 48; DB 3; Length 8920;
 Best Local Similarity 46.0%; Pred. No. 0.14;
 Matches 199; Conservative 0; Mismatches 230; Indels 2; Gaps 1;
 QY 8 ATATAGATATAGATAATGTTAAATTTAGAGAGAACAAATATGATGATTTAGGATTTAAG 67
 Db 877 AGATATATATACCTTTTAATTAACAACTTTGGATGTTAAAGAAATAAACTGTTTAGACC 818
 QY 68 AGGTCAATACAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGTTTATCAAA 127
 Db 817 TAGATTCAGAGAAATATCCCAATATATATATATATATATATATATATATATATAT 758
 QY 128 CTGGATTTCAAGATTTTAACTGTTTATTTCAAGAAATATGAATTTTGGTTAT 187
 Db 757 ATATATTTTCCCAATTTTCTTTTTTTTTTTTATACATTTAT--ATTTATGTTTAAATA 700
 QY 188 CCATATAGATCAATATAATAACTCTTCTTCACTTCGGAATCATTTGATATATTTGGAAG 247
 Db 699 TTTATAAATTTACATATACAGTTCAATTTTCATATGTAATTTTTTTTTTTCTTTTTT 640
 QY 248 ATATTATGCACAAATTTCTATAACAAATGGAGATAAGAAAGTTCAAGATTTAGAAAAA 307
 Db 639 TTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTAGTAATTTACTATTTTATAAAGTAAAGAAA 580
 QY 308 ATATTAAGGCAAGAAATAAATTAAGAGTGCATAGTAGTATACTAAGAACTTGA 367
 Db 579 AAATAAATAATCAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 520
 QY 368 AAATGATTCATTTGATTAATTTGGTGAAGAAAGTAGTGGAGCTTTTACAAGTTAGTGA 427
 Db 519 ATTGGGAGTATATATATATATATATATATATATATATATAGGATGTTTATATAGAGTGA 460
 QY 428 ACCCGATGA 437
 Db 459 AACTATATTA 450

RESULT 15

US-09-601-198-75/c
 ; Sequence 75, Application US/09601198
 ; Patent No. 6531583
 ; GENERAL INFORMATION:

; APPLICANT: Cassell, Gail H.
 ; APPLICANT: Chen, Ellison Y.
 ; APPLICANT: Glass, Jennifer S.
 ; APPLICANT: Glass, John I.
 ; APPLICANT: Heiner, Cheryl R.
 ; APPLICANT: Lefkowitz, Elliot
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
 ; FILE OF INVENTION: UREALYTICUM
 ; FILE REFERENCE: UAB-13452/22
 ; CURRENT APPLICATION NUMBER: US/09/601,198
 ; CURRENT FILING DATE: 2000-12-08
 ; PRIOR APPLICATION NUMBER: 60/073,189
 ; PRIOR FILING DATE: 1998-01-30
 ; NUMBER OF SEQ ID NOS: 181
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 75
 ; LENGTH: 5652
 ; TYPE: DNA
 ; ORGANISM: Ureaplasma urealyticum
 US-09-601-198-75

Query Match 10.6%; Score 47.2; DB 4; Length 5652;
 Best Local Similarity 47.1%; Pred. No. 0.2;
 Matches 145; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
 - QY 91 GATCAATATTTAGAGGAAAGAAATATGTTTCAAACTGGATTTCAACGATTTTAAATC 150
 Db 4542 GATCTTGATCAATTTGGTAAATCAAGTGGTAATCAAAATTAATTTACTAAATTTGATGCAACC 4483
 QY 151 ATTGGTTATATTTCAAGAAATTAATGAAATTTTGGTTATCCCATATAGATCAATATAATAAC 210
 Db 4482 ATTGGTGCCCAACAACAAAAATTAATACTACATGTTTAAGGATGAATTCACACTACTAAA 4423
 QY 211 TCCTTCTCACTTCGGATCATTTGAATAATTTGGAAGATATTATGGCACAATTTCTATA 270
 Db 4422 TCAAGTTCAATGATGAATTTTATTCAAAAATGATAATAGTATTTTATAGCATCAAGTTTACGT 4363
 QY 271 ACGAATGGAGATAAAGAAAGTTGAAGATTTGAAAAAATATTAAGGCAAGAAATAAA 330
 Db 4362 GAAATAATTACTTAAATTTTATGATTTATATGATCAAAAACAAAATTTTAAATTTT 4303
 QY 331 TTAAGAGTGATAGTAGTATATACCTTAAGAACTTTGGAATTTGATTCATTTGGATAATTG 390
 Db 4302 TATGATGTTGAATCTACATTTTAAATAAATAAACAATTTTATATATATGATATTGATGTAAT 4243
 QY 391 GTGAAGA 398
 Db 4242 GGTAAAA 4235

Search completed: February 16, 2004, 12:52:52
 Job time : 72 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 06:13:40 ; Search time 277 Seconds

(without alignments)
4356.133 Million cell updates/sec

Title: US-09-980-054a-11

Perfect score: 447

Sequence: 1 atgctagatatagataga.....accocgatgatgtgggtga 447

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneq/geneq-emb1/NA1980.DAT:*

2: /SIDSI/gcgdata/geneq/geneq-emb1/NA1981.DAT:*

3: /SIDSI/gcgdata/geneq/geneq-emb1/NA1982.DAT:*

4: /SIDSI/gcgdata/geneq/geneq-emb1/NA1983.DAT:*

5: /SIDSI/gcgdata/geneq/geneq-emb1/NA1984.DAT:*

6: /SIDSI/gcgdata/geneq/geneq-emb1/NA1985.DAT:*

7: /SIDSI/gcgdata/geneq/geneq-emb1/NA1986.DAT:*

8: /SIDSI/gcgdata/geneq/geneq-emb1/NA1987.DAT:*

9: /SIDSI/gcgdata/geneq/geneq-emb1/NA1988.DAT:*

10: /SIDSI/gcgdata/geneq/geneq-emb1/NA1989.DAT:*

11: /SIDSI/gcgdata/geneq/geneq-emb1/NA1990.DAT:*

12: /SIDSI/gcgdata/geneq/geneq-emb1/NA1991.DAT:*

13: /SIDSI/gcgdata/geneq/geneq-emb1/NA1992.DAT:*

14: /SIDSI/gcgdata/geneq/geneq-emb1/NA1993.DAT:*

15: /SIDSI/gcgdata/geneq/geneq-emb1/NA1994.DAT:*

16: /SIDSI/gcgdata/geneq/geneq-emb1/NA1995.DAT:*

17: /SIDSI/gcgdata/geneq/geneq-emb1/NA1996.DAT:*

18: /SIDSI/gcgdata/geneq/geneq-emb1/NA1997.DAT:*

19: /SIDSI/gcgdata/geneq/geneq-emb1/NA1998.DAT:*

20: /SIDSI/gcgdata/geneq/geneq-emb1/NA1999.DAT:*

21: /SIDSI/gcgdata/geneq/geneq-emb1/NA2000.DAT:*

22: /SIDSI/gcgdata/geneq/geneq-emb1/NA2001A.DAT:*

23: /SIDSI/gcgdata/geneq/geneq-emb1/NA2001B.DAT:*

24: /SIDSI/gcgdata/geneq/geneq-emb1/NA2002.DAT:*

25: /SIDSI/gcgdata/geneq/geneq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	447	100.0	447	22 AAC86750	DNA encoding a Can
2	442.2	98.9	447	22 AAS23431	Candida albicans e
3	442.2	98.9	447	24 AB231775	Candida albicans e
4	324.4	72.6	326	22 AAC86756	Probe for DNA enco
5	70.4	15.7	4985	24 ABQ75107	Anopheles gambiae
C 6	65	14.5	65	24 ABZ28869	Candida gene relat
C 7	62.6	14.0	9539	22 AAS45346	Chemically pretrea
C 8	62.6	14.0	9539	24 ABK28179	DNA transcription

9	61.2	13.7	168575	22	AAH21613	Human hypocretin r
C 10	60.2	13.5	17183	24	ABU32487	Human immune syste
C 11	59.8	13.4	6109	24	ABU32326	Human immune syste
C 12	59.8	13.4	6109	24	AAS61077	Human gene regulat
C 13	59.8	13.4	9789	17	AAAT41852	cDNA encoding plas
C 14	59.8	13.4	12237	24	ABU34358	Human immune syste
C 15	59.4	13.3	446	25	ABX38235	Bovine EST associa
C 16	59.2	13.2	11143	23	ABU12834	Drosophila melanog
C 17	59.2	13.2	12142	24	ABU33673	Human immune syste
C 18	57.8	12.9	516	25	ABX40620	Human immune syste
C 19	57.6	12.9	9539	22	AAS45347	Bovine EST associa
C 20	57.6	12.9	9539	24	ABK28180	Chemically pretrea
C 21	57.4	12.8	6641	24	ABU54335	DNA transcription
C 22	57.4	12.8	6641	24	ABU32314	Chemically treated
C 23	57.4	12.8	32392	24	ABU56203	Human immune syste
C 24	57.2	12.8	4311	21	AAAT70133	AMEV genome fragm
C 25	57	12.8	5940	21	AAAT70105	Plasmodium falcipa
C 26	56.8	12.7	17131	24	ABU33053	Human immune syste
C 27	56.6	12.7	6763	24	ABQ67067	Human angiogenesis
C 28	56.4	12.6	402	22	AA182045	Human polyucleoti
C 29	56.4	12.6	1686	16	AAQ87587	DNA encoding Leuco
C 30	56.4	12.6	6301	24	ABU32053	Human immune syste
C 31	55.8	12.5	6286	22	AAS46591	Tumour suppressor
C 32	55.6	12.4	1407	24	ABU56226	AMEV serine threo
C 33	55.6	12.4	9583	24	ABU32894	Human immune syste
C 34	55.6	12.4	19124	18	AAAT72882	Plasmodium var-7 g
C 35	55.6	12.4	19124	21	AAZ98287	Plasmodium var-7 p
C 36	55.6	12.4	50000	24	ABU56201	AMEV genome fragm
C 37	55.2	12.3	1998	21	AAAT70212	Plasmodium falcipa
C 38	55.2	12.3	4590	7	AAAG0472	Sequence encoding
C 39	55.2	12.3	6175	24	ABU33307	Human immune syste
C 40	55.2	12.3	6621	21	ABAT70188	Plasmodium falcipa
C 41	55	12.3	55	24	ABZ28979	Candida gene relat
C 42	55	12.3	80	22	AAS23676	Tetracycline promo
C 43	54.8	12.3	419	25	ABX46069	Bovine EST associa
C 44	54.8	12.3	509	24	ABQ16728	Oligonucleotide fo
C 45	54.8	12.3	509	24	ABQ16729	Oligonucleotide fo

ALIGNMENTS

RESULT 1
AAC86750
ID AAC86750 standard; DNA; 447 BP.
XX AAC86750;
XX
XX 02-APR-2001 (first entry)
XX
XX DNA encoding a Candida albicans protein CanL260.
XX
XX CADRA472; CADR489; 1CADRS527; 2CADRS527; CaPLO24; CanL260; CaDR361;
XX antifungal; fungal infection; pathogenic fungi; ss.
XX Candida albicans.
XX
XX Key Location/Qualifiers
XX CDS 1..447
XX FT /*tag= a
XX
XX WC200075305-A2.
XX
XX 14-DEC-2000.
XX
XX 08-JUN-2000; 2000WO-FR01567.
XX
XX 09-JUN-1999; 99FR-0007250.
XX
XX (HMRI) HOECHST MARION ROUSSEL.
XX
XX Lalanne J, Rocher C;

DR WPI; 2001-050024/06.
 DR P-PSDB; AAB30859.
 XX
 XX
 PT New polynucleotides from *Candida albicans* and their derived proteins,
 PT useful for diagnosis and treatment of fungal infections and for drug
 PT screening -
 XX
 XX
 PS Claim 4; Page 77-78; 89pp; French.
 XX
 CC The present sequence encodes a *Candida albicans* protein. The
 CC specification describes genes CaDRA472, CaDR489, CaDR527, CaDR527,
 CC CaFLO24, CaNL260, and CaNR361. These genes are essential for
 CC survival, and so are good targets for antifungal agents. The *Candida*
 CC *albicans* genes and their derived proteins are used to screen compounds
 CC for the ability to inhibit the activity of the protein, i.e. for
 CC antifungal activity. The proteins are also used to generate a protective
 CC antibody response against fungal infections in mammals. The *Candida*
 CC *albicans* proteins and genes, and their antibodies, are used for
 CC diagnosing fungal infections, specifically *C. albicans* (in standard
 CC amplification, hybridisation or immunological assays, and for studying
 CC pathogenic fungi.
 XX
 XX Sequence 447 BP; 192 A; 38 C; 82 G; 135 T; 0 other;
 CC
 CC Query Match 100.0%; Score 447; DB 22; Length 447;
 CC Best Local Similarity 100.0%; Pred. No. 1.2e-70;
 CC Matches 447; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAGACAAATATGAATTAGGA 60
 Db 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAGACAAATATGAATTAGGA 60
 QY 61 TTTAAGAGGTCAAATACAGGACAAAGATCAATTTTAGAGGAGAAAGATATGCT 120
 Db 61 TTTAAGAGGTCAAATACAGGACAAAGATCAATTTTAGAGGAGAAAGATATGCT 120
 QY 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGTTTATTTCAAGATTAATGAAATTT 180
 Db 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGTTTATTTCAAGATTAATGAAATTT 180
 QY 181 TGGTTATCCCATATAGATCAATATTAATCACTCTTCACTTCGGAATCATTTGAATAAT 240
 Db 181 TGGTTATCCCATATAGATCAATATTAATCACTCTTCACTTCGGAATCATTTGAATAAT 240
 QY 241 TTGGAAGATATATGGCCAAATTTCTATACGAATGAGATAGAGATTAAGAGTTGAAGATTAT 300
 Db 241 TTGGAAGATATATGGCCAAATTTCTATACGAATGAGATAGAGATTAAGAGTTGAAGATTAT 300
 QY 301 GAAAAAATATTAAGAGGCAAGAAATTAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
 Db 301 GAAAAAATATTAAGAGGCAAGAAATTAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
 QY 361 ACTTGGAATATGATTCATTCGTAATTTGGTGAAGAGTAGGTGGAACTTTTCAAGTT 420
 Db 361 ACTTGGAATATGATTCATTCGTAATTTGGTGAAGAGTAGGTGGAACTTTTCAAGTT 420
 QY 421 AGTGAACCCCGATGATATGTTGTTGA 447
 Db 421 AGTGAACCCCGATGATATGTTGTTGA 447
 RESULT 2
 AAS23431
 ID AAS23431 standard; DNA; 447 BP.
 XX
 AC AAS23431;
 XX
 XX
 DT 04-DEC-2001 (first entry)
 XX
 DE -*Candida albicans* essential gene CaYNL260C.
 XX
 KW Gene identification; essential gene; GRACE; pathogenic fungus;
 KW gene replacement and conditional expression; fungal infection; ds.

XX
 OS
 XX
 PN
 XX
 PD
 XX
 XX
 PP
 XX
 XX
 PR
 XX
 PA
 XX
 PI
 XX
 DR
 XX
 DR
 XX
 PT
 XX
 PT
 XX
 PS
 XX

Candida albicans.
 WO200160975-A2.
 23-AUG-2001.
 20-FEB-2001; 2001WO-US05551.
 18-FEB-2000; 2000US-0183534.
 (ELIT-) ELITRA PHARM INC.
 Roemer T, Jiang B, Boone C, Bussey H;
 WPI; 2001-489080/53.
 P-PSDB; AAU15103.

Identifying genes essential to fungal metabolisms and identifying
 potential therapeutic agents that target these genes -
 Claim 22; Page 167; 324pp; English.

The present invention relates to novel methods for constructing fungal
 strains useful for identification and validation of gene products as
 targets for therapeutic agents, for creating a collection of identified
 essential genes, and screening assays for the discovery of new drugs.
 The invention provides the GRACE (gene replacement and conditional
 expression) method for the construction of mutant organisms referred to
 as GRACE strains of the organism. The invention can be applied to any
 organism, particularly a pathogenic fungus e.g. *Candida albicans*,
Aspergillus fumigatus and *Cryptococcus neoformans*. The methods are
 useful to identify agents that may be used in the treatment of fungal
 infections. AAS23381-AAS23442 represent *C. albicans* essential genes.

Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;

Query Match 98.9%; Score 442.2; DB 22; Length 447;
 Best Local Similarity 99.3%; Pred. No. 8.7e-70;
 Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGTCAGATATAGATATAGATATGTTTAAATTTAGAGAGAGACAAATATGAATTAGGA 60
 Db 1 ATGTCAGATATAGATATAGATATGTTTAAATTTAGAGAGAGACAAATATGAATTAGGA 60
 QY 61 TTTAAGAGGTCAAATACAGGACAAAGATCAATTTTAGAGAGAGAAAGATATGCT 120
 Db 61 TTTAAGAGGTCAAATACAGGACAAAGATCAATTTTAGAGAGAGAAAGATATGCT 120
 QY 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGTTTATTTCAAGATTAATGAAATTT 180
 Db 121 TATCAAACTGGATTTCAACGATTTTAAATCATTTGTTTATTTCAAGATTAATGAAATTT 180
 QY 181 TGGTTATCCCATATAGATCAATATTAATCACTCTTCACTTCGGAATCATTTGAATAAT 240
 Db 181 TGGTTATCCCATATAGATCAATATTAATCACTCTTCACTTCGGAATCATTTGAATAAT 240
 QY 241 TTGGAAGATATATGGCCAAATTTCTATACGAATGAGATAGAGATTAAGAGTTGAAGATTAT 300
 Db 241 TTGGAAGATATATGGCCAAATTTCTATACGAATGAGATAGAGATTAAGAGTTGAAGATTAT 300
 QY 301 GAAAAAATATTAAGAGGCAAGAAATTAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
 Db 301 GAAAAAATATTAAGAGGCAAGAAATTAATTAAGAGTGATAGCTAGTATAACTAAAGAA 360
 QY 361 ACTTGGAATATGATTCATTCGTAATTTGGTGAAGAGTAGGTGGAACTTTTCAAGTT 420
 Db 361 ACTTGGAATATGATTCATTCGTAATTTGGTGAAGAGTAGGTGGAACTTTTCAAGTT 420
 QY 421 AGTGAACCCCGATGATATGTTGTTGA 447
 Db 421 AGTGAACCCCGATGATATGTTGTTGA 447

RESULT 3
 ABZ31775
 ID ABZ31775: standard; DNA; 447 BP.
 XX AC ABZ31775;
 XX DT 30-JAN-2003 (first entry)
 XX DE Candida albicans essential gene SEQ ID NO 6062.
 XX KW Fungus; Yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 XX KW signal transduction; DNA replication; cell division; growth;
 XX KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
 XX CS Candida albicans.
 XX PN WO200253728-A2.
 XX PD 11-JUL-2002.
 XX PF 26-DEC-2001; 2001WO-US49486.
 XX PR 29-DEC-2000; 2000US-259128P.
 XX PR 20-FEB-2001; 2001US-0792024.
 XX PR 22-AUG-2001; 2001US-314050P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 XX WPI; 2002-566694/60.
 XX DR P-PSDB; ABP73225.
 XX
 PT Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele
 PT of a gene and placing other allele of the gene under conditional
 PT expression
 XX
 PS Claim 37; SEQ ID NO 6062; 167pp + Sequence Listing; English.
 CC The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of an
 CC essential Candida albicans gene used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 SQ Sequence 447 BP; 193 A; 38 C; 80 G; 136 T; 0 other;
 Query Match 98.9%; Score 442.2; DB 24; Length 447;
 Best Local Similarity 99.3%; Pred. No. 8.7e-70;
 Matches 444; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 CC The Candida albicans genes and their derived proteins are used to screen

QY 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGAATTAGGA 60
 DB 1 ATGTCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGAATTAGGA 60
 QY 61 TTTAAAGAGGTCAAATACAGGAGCAAAAGATCAATATTTAGAGAGAAAGATATGGT 120
 DB 61 TTTAAAGAGGTCAAATACAGGAGCAAAAGATCAATATTTAGAGAGAAAGATATGGT 120
 QY 121 TATCAAACTGGATTTCAACGATTTTAAATCAATGTTTATTTCAAGAAATTAATGAAATTT 180
 DB 121 TATCAAACTGGATTTCAACGATTTTAAATCAATGTTTATTTCAAGAAATTAATGAAATTT 180
 QY 181 TGGTTATCCCATATAGATCAATATAAATCTTCTTCACTTCGGAATCAATTTGAATAAT 240
 DB 181 TGGTTATCCCATATAGATCAATATAAATCTTCTTCACTTCGGAATCAATTTGAATAAT 240
 QY 241 TTGGAAGATATATGCGCACAAATTTCTATAACGAATCGAGATAAAGAGTTGAAGATTAT 300
 DB 241 TTGGAAGATATATGCGCACAAATTTCTATAACGAATCGAGATAAAGAGTTGAAGATTAT 300
 QY 301 GAAAAAATATTTAAAAAGCGCAAGAAATAAATTAAGAGTGTAGCTAGTATAAAGAA 360
 DB 301 GAAAAAATATTTAAAAAGCGCAAGAAATAAATTAAGAGTGTAGCTAGTATAAAGAA 360
 QY 361 ACTTGGAAAATTTGATTCATTTGGATTAATTTGGTGAAGAAAGTAGGTGAACTTTACAAGTT 420
 DB 361 ACTTGGAAAATTTGATTCATTTGGATTAATTTGGTGAAGAAAGTAGGTGAACTTTACAAGTT 420
 QY 421 AGTGAAAAACCCCGATGATATGTGTGA 447
 DB 421 AGTGAAAAACCCCGATGATATGTGTGA 447

RESULT 4
 AAC86756
 ID AAC86756 standard; DNA; 326 BP.
 XX AC AAC86756;
 XX DT 02-APR-2001 (first entry)
 XX DE Probe for DNA encoding a Candida albicans protein CanL260.
 XX KW CaDR4472; CaDR489; 1CaDR527; 2CaDR527; CaPLO24; CanL260; CaDR361;
 XX antifungal; fungal infection; pathogenic fungi; probe; ss.
 XX OS Candida albicans.
 XX PN WO200075305-A2.
 XX PD 14-DEC-2000.
 XX PF 08-JUN-2000; 2000WO-FR01567.
 XX PR 09-JUN-1999; 99FR-0007250.
 XX (HMRI) HOECHST MARION ROUSSEL.
 XX Lalanne J, Rocher C;
 XX WPI; 2001-050024/06.
 XX New polynucleotides from Candida albicans and their derived proteins,
 XX useful for diagnosis and treatment of fungal infections and for drug
 XX screening
 XX Example 5; Page 85-86; 89pp; French.
 XX The present sequence represents a probe for DNA encoding a Candida
 XX albicans protein. The specification describes genes CaDR4472, CaDR489,
 XX 1CaDR527, 2CaDR527, CaPLO24, CanL260, and CaDR361. These genes are
 XX essential for survival, and so are good targets for antifungal agents.
 XX The Candida albicans genes and their derived proteins are used to screen

KW Fungus; yeast; tetracycline; promoter; GRACE strain; biosynthesis;
 KW signal transduction; DNA replication; cell division; growth;
 KW proliferation; Candida albicans; fungicide; antifungal; PCR; primer; ss.
 XX
 OS Candida albicans.
 XX
 PN WO200253728-A2.
 XX
 XX 11-JUL-2002.
 XX
 XX 26-DEC-2001; 2001WO-US49486.
 XX
 XX 29-DEC-2000; 2000US-259128P.
 PR 20-FEB-2001; 2001US-0792024.
 PR 22-AUG-2001; 2001US-314050P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 PA
 XX
 XX Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
 PI
 XX
 XX WPI; 2002-566694/60.
 DR
 XX

XX Constructing strains for identifying gene products as effective targets
 PT for therapeutic intervention, by inactivating in the strain one allele
 PT of a gene and placing other allele of the gene under conditional
 PT expression
 PT
 XX
 XX Claim 76; SEQ ID NO 2884; 167pp + Sequence Listing; English.

XX The invention relates to constructing (M1) a strain of diploid fungal
 CC cells in which both alleles of a gene are modified, comprising modifying
 CC one allele by insertion or replacement by a cassette having an
 CC expressible selectable marker and modifying other allele by
 CC recombination, of a promoter replacement fragment with a heterologous
 CC promoter, so that expression of the second allele is regulated by the
 CC promoter. (M1) is useful for constructing a strain of diploid fungal
 CC cells in which both alleles of a gene are modified. The diploid fungal
 CC cells having both alleles modified are useful for identifying a gene that
 CC is essential to the survival or growth of a fungus, a gene that
 CC contributes to the virulence and/or pathogenicity of a fungus, a gene
 CC that contributes to the resistance of a diploid fungus to an antifungal
 CC agent, an antifungal agent that inhibits the growth of a diploid fungus
 CC and for identifying a therapeutic agent for treatment of a mammalian
 CC disease. (M1) is useful for identifying a compound which modulates the
 CC activity of a gene product, preferably enzymatic activity, carbon
 CC compound catabolism, biosynthetic, transporter, transcriptional,
 CC translational, signal transduction, DNA replication and cell division
 CC activity. The method is useful for identifying a compound having the
 CC ability to inhibit growth or proliferation of C. albicans cells and for
 CC treating infection by C. albicans. The present sequence is that of a PCR
 CC primer used in the method of the invention.
 CC Note: The sequence data for this patent is not represented in the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX

SQ Sequence 65 BP; 20 A; 11 C; 2 G; 32 T; 0 other;

Query Match 14.5%; Score 65; DB 24; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.005;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGATGATTT 63
 DB 65 TCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGATGATTT 6

QY 64 AAAGA 68

DB 5 AAAGA 1

RESULT 7
 AAS45346/C
 ID AAS45346 standard; DNA; 9539 BP.

XX AAS45346;
 AC
 XX 18-DEC-2001 (first entry)
 DT
 XX
 DE Chemically pretreated genomic DNA associated with cell cycle #26.
 XX
 KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
 KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
 KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
 KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
 KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
 KW PCR primer.
 XX
 OS Homo sapiens.
 XX
 PN WO200168911-A2.
 XX
 XX 20-SEP-2001.
 PD
 XX
 XX 15-MAR-2001; 2001WO-EP02945.
 PP
 XX 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2001-602751/68.
 DR
 XX
 XX Designing primers and probes for analysing diseases associated with
 PT cytosine methylation state e.g. arthritis, cancer, aging,
 PT arteriosclerosis comprising fragments of chemically modified genes
 PT associated with cell cycle -
 PT
 XX Claim 1; SEQ ID No 51; 28pp; English.

XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
 CC molecules associated with the cell cycle and specific PCR primers of the
 CC invention. The sequences are useful for detecting the methylation state
 CC of all CpG dinucleotides in a sequence and therefore for analysing
 CC associated diseases. By analysing cytosine methylations in the pretreated
 CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
 CC of existing diseases or the predisposition to specific diseases can be
 CC ascertained. The parameters may be compared to another set of genetic
 CC and/or epigenetic parameters, the differences serving as basis for
 CC diagnosis and/or prognosis events which are disadvantageous to patients.
 CC The sequences of the invention are useful for the diagnosis and therapy
 CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
 CC aging, glomerular disease, Lewy body disease, arthritis,
 CC arteriosclerosis, solid tumours and cancers.
 XX

SQ Sequence 9539 BP; 2665 A; 121 C; 2515 G; 4238 T; 0 other;

Query Match 14.0%; Score 62.6; DB 22; Length 9539;
 Best Local Similarity 47.8%; Pred. No. 0.014;
 Matches 182; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 4 TCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGATGATTT 63
 DB 6192 TCAGATATAGATATAGATATATTAATTTAGAGAGAACCAATATGATGATTT 6133

QY 64 AAAGAAGTCAATATCAAGGAACAAAGATCAATATTTAGAGAGAAAGATATGTTAT 123

DB 6132 AAAACAAAGAACCACTTAAAAATTCAAAAACACATTTTAACTCAAAATAATATATAA 6073

QY 124 CAACTGGATTTCAAGGATTTTATCATCTGTTATTTCAAGATTTATGAAATTTGG 183

DB 6072 AACAAACGAATTTAAAAATTCGAAATTTAAAAATTCAAAAATTTAAAAATA 6013

Qy	184	TTATCCCATATAGATCAATATAATAA	CTCTTCTCACTTCGGAATCA	TTTGAATAATTG	243
Db	6012	ATATAAAAATCAAAATCTTAAAAAT	CTTAAAAATCTAAAAAT	TTTTCTTAA	5953
Qy	244	GAAGATATTATGGCACAAATTTCTAT	AACGAAATGGAGATAAAGAA	GTGTGAAGATATATGAA	303
Db	5952	AAAAAAAATAAAATTTAAAAATCAAA	ATCTATTTTAAACAAACAAATATA	TAACCTAATAAA	5893
Qy	304	AAAAAATATTTAAAAAGCGCAAGAA	ATAAATAAAGTGTAGCTAGTATAT	ACTTAACAGAACT	363
Db	5892	AAAAAATTTTAAAAAACCAATAAAT	TAAATTAATAAAAAAACTTTAAAC	TAATAAAAAAATT	5893
Qy	364	TGGAAAATTCATTCATTGGAT			384
Db	5832	CAAAAAAACCAATAAAATATAT			5812

RESULT 8	
ABK28179/c	
ID	ABK28179 standard; DNA; 9539 BP.
XX	
XX	ABK28179;
XX	
XX	
XX	23-APR-2002 (first entry)
XX	
XX	DNA transcription associated genomic DNA #27.
XX	
KW	DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW	PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW	single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW	viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW	immunological disorder; Werner syndrome; developmental disorder;
KW	psoriasis; Kieleg's syndrome; neurological disorder; erythropoiesis;
KW	neurodegenerative disorder; Wardenburg syndrome; Niemann-Pick disease;
KW	myelocystic syndrome; myocardial infarction; hypertension; arthritis
KW	angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW	polyglutamine disorder; solid tumour.

XX	Unidentified.
XX	OS
XX	XX
XX	WO200192565-A2.
XX	PN
XX	XX
XX	06-DEC-2001.
XX	PD
XX	XX
XX	06-APR-2001, 2001WO-EP03973.
XX	PF
XX	XX
XX	06-APR-2000; 2000DE-1019058.
XX	07-APR-2000; 2000DE-1019173.
XX	30-JUN-2000; 2000DE-1032529.
XX	01-SEP-2000; 2000DE-1043826.
XX	XX
XX	(EPIG-) EPIGENOMICS AG.
XX	PA
XX	XX
XX	Olek A, Piepenbrock C, Berlin K;
XX	XX
XX	WPI; 2002-090046/12.
XX	DR
XX	XX
XX	PT New nucleic acids or oligomers, useful for diagnosing or treating
PT	diseases associated with DNA transcription, e.g. immunological
PT	disorders, Werner syndrome, psoriasis, myocardial infarction, solid
PT	tumours or cancer
PT	PT

PS
xx
Claim 1; SEQ ID No 53; 32pp; English.
The invention relates to a nucleic acid, which comprises a segment of the chemically pretreated DNA of genes associated with DNA transcription from one of 346 sequences, and an oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical to the chemically pretreated DNA of genes associated with DNA transcription. The set of oligomer probes are useful for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA. The nucleic acids are useful for

diagnosing or treating diseases associated with DNA transcription (particularly with the methylation status), e.g. adenosine deaminase deficiency, viral infection, retroviral infection, Sezary syndrome, haematological disorders, immunological disorders, Werner syndrome, tuberulosis, developmental disorders, psoriasis, Rieger's syndrome, neurological disorders, neurodegenerative disorders, Waardenburg syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial infarction, hypertension, angiogenesis, erythroiposis, congenital heart disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours or cancer. Sequences ABK28127-ABK28472 represent DNA transcription associated genomic DNA molecules of the invention.

Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format directly from the European Patent Office.

Query Match	14.0%;	Score 62.6;	DB 24;	Length 9539;
Best Local Similarity	47.8%;	Pred. No. 0.014;		
Matches 182;	Conservative 0;	Mismatches 199;	Indels 0;	Gaps 0;
Qy	4	TCAGATATAGATATAGATATATGATTAAATTTAGAAGAGAACAAATATGAATTAGGATTT	63	
Db	6192	TCAAAAAATTTAAATATTAACCATTAATTTTATAAAAAACAATATATACCGAATTAT	6133	
Qy	64	AAAGAAGGTCAAATACAGGAACAAAGATCAATATTTTGAAGGAAAGAAATATGGTTAT	123	
Db	6132	AAAAACAAAACGAACTTAAAAATTTCAAAAACACATTTTAAACTCAAAAAATAAATATATAA	6073	
Qy	124	CAAACTGGATTTTCAACGATTTTTTAATCATTTGGTTATATTCACGAATTAACGAAATTTGG	183	
Db	6072	AACAAACGAATTTAAAAATCGAAACGAAATTTAAAAATCAAAAATAAAATTTAAAAAATA	6013	
Qy	184	TTATCCCATATAGATCAATATATATAACTCTCTTCTACCTTCGGAATCATTTGGAATAATTTG	243	
Db	6012	ATATAAAATCAAAATTTCTAAAGATCATAAAAAATCTAAAAAATTTTCTCTAA	5953	
Qy	244	GAGATATTTATGGCACAAAATTTCTATACGAATCGAGATTAAGAACTTGAAGTTATGAA	303	
Db	5952	AAAAAAAATAAAATTTAAAAATCAAAAATCTATTTTAAACAAAACAAAATATACTAATAA	5893	
Qy	304	AAAAATATTAAAAANGCAAGAAATAAATTAAGAGTCATAGCTAGTATAACTAAAGAAACT	363	
Db	5892	AAAAAATTTAAAAACCAAAATAAANTTAATAAAAAAAAACCTTAAACATAAAAAAT	5833	
Qy	364	TGGAAAAATTGATTCATTGGAT	384	
Db	5832	CAAAAAACAATAAAATATAT	5812	

RESULT 9	
AAH21613	
ID	AAH21613 standard; DNA; 168575 BP.
XX	
XX	AAH21613;
AC	
XX	
DT	10-AUG-2001 (first entry)
XX	
XX	Human hypocretin receptor 2 (HCRTR2) gene SEQ ID NO:1.
DE	
DE	
XX	
XX	Human; narcolepsy; hypocretin receptor 2; orexin receptor 2; HCRTR2;
KW	diagnosis; ds.
KW	
XX	
XX	Homo sapiens.
OS	
XX	
XX	WO200130991-A2.
PN	
XX	
XX	
PD	03-MAY-2001.
XX	
XX	22-AUG-2000; 2000WO-US23021.
PF	
XX	
XX	25-OCT-1999; 99US-0426290.
PR	
XX	

Db 3935 TATATAAATATACCAACCAAAAAAACAATTAATAAATAAATATATATCAATCAATA 3876
 QY 360 AACTTGGAAAA 370
 Db 3875 ATCCAATAAAA 3865

RESULT 11

ABL32326/c
 ID ABL32326 standard; DNA; 6109 BP.

AC ABL32326;
 XX
 DT 26-MAR-2002 (first entry)
 XX

DE Human immune system associated gene SEQ ID NO: 299.
 XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosolic; nontropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX

OS Homo sapiens.
 XX

PN WO200200928-A2.
 XX

XX 03-JAN-2002.
 PD

XX 02-JUL-2001; 2001WO-EP07537.
 PF

XX 30-JUN-2000; 2000DE-1032529.
 PR

XX 01-SEP-2000; 2000DE-1043826.
 PR

XX (EPIG-) EPIGENOMICS AG.
 PA

XX Olek A, Piepenbrock C, Berlin K;
 PI

XX WPI; 2002-130909/17.
 DR

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX

PS Claim 1; SEQ ID NO 299; 32pp + Sequence Listing; German.
 XX

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX

XX Sequence 6109 BP; 1232 A; 195 C; 1734 G; 2944 T; 4 other;
 SQ

Query Match 13.4%; Score 59.8; DB 24; Length 6109;
 Best Local Similarity 46.3%; Pred. No. 0.043;
 Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 6 AGATATAGATATAGTAATCTTATTAATTTAGAGAGACATATGAATTAGGATTAA 65
 Db 2081 AATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2022

QY 66 AGAGGTCAATCAAGGACAAAAGATCAATATTTAGAGGAAAAAGATATGGTTATCA 125

Db 2021 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 1962

QY 126 AACTGGATTCAACGATTTTAAATCATCTGGTTATATTCAGGAATTAATGAATTTGGTT 185

Db 1961 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 1902
 QY 186 ATCCCATATAGATCATATAATAAATCTCTTCTTCACCTTCGGAATCATTTGGAATTTGGA 245
 Db 1901 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 1842
 QY 246 AGATATTTATGGCACAAATTTCTTATAACGAATCGAGATAAAGAGTTTGAAGATTATGAAA 305
 Db 1841 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 1782
 QY 306 AATATTTAAAGGCAAGAAATAAATTAAGATGATAGCTAGTATTAACATAAGAACTTG 365
 Db 1781 ATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 1722
 QY 366 GAAATTTGATTCATTGGATAAATTTGGTGAAGAACTAGTGGAACTTTTCAAGTTAGTGA 425
 Db 1721 ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAAT 1662
 QY 426 AAA 428
 Db 1661 ATA 1659

RESULT 12

AAS61077/c

ID AAS61077 standard; DNA; 6109 BP.

XX
 AC AAS61077;
 XX

DT 29-JAN-2002 (first entry)
 XX

DE Human gene regulation-associated gene oligonucleotide #32.
 XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
 KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
 KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
 KW renal disease; Preclampsia; cardiac allograft vascular disease;
 KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
 KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
 KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
 XX

OS Homo sapiens.
 XX

PN WO200177375-A2.
 XX

XX 18-OCT-2001.
 PD

XX 06-APR-2001; 2001WO-EP03969.
 PF

XX 06-APR-2000; 2000DE-1019058.
 PR

XX 07-APR-2000; 2000DE-1019173.
 PR

XX 30-JUN-2000; 2000DE-1032529.
 PR

XX 01-SEP-2000; 2000DE-1043826.
 PR

XX (EPIG-) EPIGENOMICS AG.
 PA

PI Olek A, Piepenbrock C, Berlin K;
 PI

XX WPI; 2002-017470/02.
 DR

XX New nucleic acid sequences from chemically modified genes associated
 PT with gene regulation, useful for analysing cytosine methylations for
 PT diagnosis and therapy of diseases e.g. severe combined immunodeficiency
 PT disease -
 XX

PS Claim 1; SEQ ID NO 33; 26pp; English.
 XX

CC The invention relates to 224 nucleic acid sequences comprising at least
 CC 18 bases of a chemically pretreated gene associated with gene regulation
 CC selected from 43 known genes (or complementary sequences). The
 CC chemical pretreatment converts cytosine bases unmethylated at the
 CC 5-position to uracil or another base with hybridisation behaviour

CC PfEMP1 protein of the MC type of Plasmodium falciparum. An
 CC alternative, truncated version of the coding sequence (a cDNA clone)
 CC is given in AAT41853.
 XX
 SQ Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T; 0 other;

Query Match 13.4%; Score 59.8; DB 17; Length 9789;
 Best Local Similarity 47.1%; Pred. No. 0.043;
 Matches 184; Conservative 0; Mismatches 207; Indels 0; Gaps 0;

QY 6 AGATATAGATATAGATATATGTTAAATTTAGAGAGAGACAAATATGCAATTAGGATTAA 65
 DB 7888 AAAGAAATATATAAAAAATTTATTAAATAATAAATAAAAAATTCATATAAAAAA 7947

QY 66 AGAGCTCAATATACAGGAGCAAAAGATCAATATTTAGAGGAGAAAGAAATGTTATCA 125
 DB 7948 AGAAATGAAATTTTATATAAAAAAATTTATTAATAAATAATGATTATATAAAAAA 8007

QY 126 AACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAAGAATTAATGAAATTTTGGTT 185
 DB 8008 ATTATTAGAAATAAATAAATAAAAAATTTATTAAATAAATAAAGAAAAAATG 8067

QY 186 ATCCCATATAGATCAATATATAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 245
 DB 8068 TTAAAAAT 8127

QY 246 AGATATTTATGCACAAATTTCTATAACGAATGGAGATTAAGAGTTGCAAGTTATGAAA 305
 DB 8128 TTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 8187

QY 306 AAATATTAAAAAGGCAAGAAATAAATTAAAGAGTGTAGTAGTATTAACATAAGAACTTG 365
 DB 8188 AATAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8247

QY 366 GAAATTCATTCATTTGGATTTTGGTGAA 396
 DB 8248 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8278

RESULT 14
 ABL34358/c
 ID ABL34358 standard; DNA; 12237 BP.
 XX
 AC ABL34358;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 2331.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cyostatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritis; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 2331; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 12237 BP; 3050 A; 207 C; 3024 G; 5956 T; 0 other;

Query Match 13.4%; Score 59.8; DB 24; Length 12237;
 Best Local Similarity 46.3%; Pred. No. 0.043;
 Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 6 AGATATAGATATAGATATATGTTAAATTTAGAGAGAGACAAATATGCAATTAGGATTAA 65
 DB 6985 AAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6926

QY 66 AGAGCTCAATATACAGGAGCAAAAGATCAATATTTAGAGGAGAAAGATATGTTATCA 125
 DB 6925 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6866

QY 126 AACTGGATTTCAACGATTTTAAATCATTTGGTTATATTCAAGAATTTAAATGAAATTTTGGTT 185
 DB 6865 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6806

QY 186 ATCCCATATAGATCAATATATAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 245
 DB 6805 AAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6746

QY 246 AGATATTATGGCACAATTTCTATAACGAATGGAGATTAAGAGTTTGAAGTTATGAAA 305
 DB 6745 TAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6686

QY 306 AAATATTAAAAAGGCAAGAAATAAATTAAAGAGTGTAGTAGTATTAACATAAGAACTTG 365
 DB 6685 TAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6626

QY 366 GAAATTCATTCATTTGGGATAATTTGTCGAGAGAGTAGTGGAACTTTTACAAAGTTAGTGA 425
 DB 6625 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6566

QY 426 AAA 428
 DB 6565 ATA 6563

RESULT 15
 ABX38235/c
 ID ABX38235 standard; cDNA; 446 BP.
 XX
 AC ABX38235;
 XX
 DT 20-FEB-2003 (first entry)
 XX
 DE Bovine EST associated with lactation/muscle/fat deposition #3400.
 XX
 KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
 KW muscle deposition; fat deposition; genome mapping; gene identification;
 KW gene analysis; cattle breeding.
 XX
 OS Bos Taurus.
 XX
 PN US2002137139-A1.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2004, 10:54:01 ; Search time 2562 Seconds
(without alignments)
4240.477 Million cell updates/sec

Title: US-09-980-054a-11

Perfect score: 447

Sequence: 1 atgtcagatatagataga.....accocgatgatgtggtga 447

Scoring tables: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_estbam:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	208.2	46.6	1059	29	CNS07EAO
C 2	115.6	25.9	902	29	CNS06VIL
C 3	80.8	18.1	994	13	AL417187 T3 end of
C 4	78.2	17.5	1101	29	AL414650 BX414650
					AL060732 Drosophil

5	76.4	17.1	1201	9	AL565455
C 6	75.6	16.9	1200	13	BX415878
C 7	74.4	16.6	1101	29	CNS00EVL
C 8	74	16.6	1201	13	BX462207
C 9	72.4	16.2	725	28	BH180166
C 10	72.4	16.2	1201	13	BX439779
C 11	71.2	15.9	1201	13	BX335216
C 12	71	15.9	1201	29	CNS0167M
C 13	70.6	15.8	1101	29	CNS00EA9
C 14	70.2	15.7	997	14	CD049644
C 15	70	15.7	1001	29	CNS0155H
C 16	69.8	15.6	942	29	CNS018GS
C 17	69.8	15.6	1201	9	AL543368
C 18	69	15.4	639	29	CNS038CX
C 19	69	15.4	1126	13	BX446388
C 20	69	15.4	1164	29	CC218891
C 21	68.6	15.3	1044	13	BX415231
C 22	68.6	15.3	1139	28	AQ897537
C 23	68.6	15.3	1201	9	AL532464
C 24	68	15.2	457	12	BJ333457
C 25	68	15.2	885	29	AG136199
C 26	67.8	15.2	455	12	BJ335466
C 27	67.8	15.2	486	12	BJ394994
C 28	67.6	15.1	1056	13	BX415058
C 29	67.4	15.1	1201	13	BX355654
C 30	67	15.0	945	29	CNS04DOK
C 31	67	15.0	1101	29	CNS015LI
C 32	66.8	14.9	388	29	CNS039KI
C 33	66.8	14.9	1201	9	AL547143
C 34	66.6	14.9	865	28	BH180441
C 35	66.6	14.9	865	29	CNS07MAN
C 36	66.6	14.9	928	29	CNS00DKY
C 37	66.6	14.9	953	13	BX331505
C 38	66.6	14.9	1169	29	CNS06KHQ
C 39	66.6	14.9	1201	13	BX421282
C 40	66.4	14.9	903	29	CNS02WVM
C 41	66.4	14.9	916	28	BH130447
C 42	66.4	14.9	1101	29	CNS0021J
C 43	66.4	14.9	1181	13	BX387370
C 44	66.4	14.9	1201	9	AL532464
C 45	66.2	14.8	768	29	CNS01VSE

ALIGNMENTS

RESULT 1	CNS07EAO	1059 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T3 end of clone XBD0AA002E07 of library XBD0AA from strain CBS 94				
DEFINITION	of Candida tropicalis, genomic survey sequence.				
ACCESSION	AL441526				
VERSION	AL441526.1	GI:12224752			
KEYWORDS	GSS.				
SOURCE	Candida tropicalis				
ORGANISM	Candida tropicalis				
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.				
AUTHORS	1 (bases 1 to 1059) Souciet,J.L., Aigle,M., Artiguenave,P., Blandin,G., Bouciet,J.L., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Jouvei,M., Wincker,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies				
JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)				
MEDLINE	20584711				
PUBMED	11152876				
REFERENCE	2 (bases 1 to 1059)				
AUTHORS	Blandin,G., Ozier-Kalogeropoulos,O., Wincker,P., Artiguenave,F. and Dujon,B.				

TITLE
Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis

JOURNAL
FEBS Lett. 487 (1), 91-94 (2000)

MEDLINE
20584726

PUBMED
11152891

REFERENCE
3 (bases 1 to 1059)

AUTHORS
Direct Submission

TITLE
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

JOURNAL
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
Location/Qualifiers
1..1059
/organism="Candida tropicalis"
/mol_type="genomic DNA"
/strain="CBS 94"
/db_xref="taxon:5482"
/clone="XBDOAA002E07"
/clone_lib="XBDOAA"
/note="end : T3"
complement(<26..>346)
/note="similar to Saccharomyces cerevisiae ORF YNL260c [hypothetical protein]"
/evidence=not_experimental

BASE COUNT
362 a 149 c 198 g 348 t 2 others

ORIGIN
Query Match 46.6%; Score 208.2; DB 29; Length 1059;
Best Local Similarity 72.8%; Pred. No. 1.3e-16;
Matches 267; Conservative 1; Mismatches 99; Indels 0; Gaps 0;

QY 1 ATGTCAGATATAGATATAGATATATGATTAATTTAGAAAGAACCAATATGATTAAGGA 60
DB 367 ATGAGCGATTTTACATATATCAAGTATTTAACTAGAGAGAACCAATATATTTAGGA 308
QY 61 TTTAAGAGGCTCAATACAGGACCAAAAGATCAATTTAGAGAAAGAAATATGTT 120
DB 307 TTTAAGGAGGACAGGACACTCTACCAAGAACCAATCTTAGAAGGAAACCAATATGTT 248
QY 121 TATCAAACTGGATTTCAACGATTTTATATCAATTCGTTATATTTCAAGATTTAATGAAATTT 180
DB 247 TATCAAACTGGTTTCAAGGTTTATTTATTTGGCTATATTCAGGATTTGGTAGGGAA 188
QY 181 TGGTATCCCATATAGATCAATATAATCACTCTTCTTCCTTCAGGATCAATTTGAATAT 240
DB 187 TGGTTGGATATCTAGACCAATATAATGCTTTCCAAATCACTTCAAGGTCATATAAATCAA 128
QY 241 TTGGAAGATATATGGGACAAATTTCTATAACGAATCGAGATTAAGATTAAGATTAAT 300
DB 127 CTAAGTGAATTAATCACTGACATACCATTCGCAATGAGATGAGATTTGAATATC 68
QY 301 GAAAAAATATATAAAGGCGAAGAAATTAATTAAGAGTGTAGCTAGTATATACTAAAGAA 360
DB 67 GAAAAAATCATCAAAAGGCGAAGAAATTAATTAAGAGTGTAGCTAGTATATACTAAAGAA 8
QY 361 ACTTTGGA 367
DB 7 AATTGGA 1

RESULT 2
CNS06VIL 902 bp DNA linear GSS 06-JUL-2001
LOCUS

DEFINITION
T3 end of clone AX0AA021C12 of library AX0AA from strain CBS 7064 of Pichia farinosa, genomic survey sequence.

ACCESSION
AL417187

VERSION
AL417187.1 GI:12198580

KEYWORDS
GSS.

SOURCE
Pichia farinosa

ORGANISM
Pichia farinosa

REFERENCE
1 (bases 1 to 902)
Boulet J.L., Aigle M., Artiguenave F., Blandin G., Bolotin-Fukuhara M., Bon E., Brotier P., Casaregola S., de Montigny J., Dujon B., Durans P., Lepingle A., Lorente B., Valperly A., Neuvéglise C., Ozier-Kalogeropoulos O., Potier S., Saurin W., Tekala F., Toffano-Nioche C., Weslowski-Louvel M., Wincker P. and Weissenbach J.

AUTHORS
Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

TITLE
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL
MEDLINE
20584711

PUBMED
11152876

REFERENCE
2 (bases 1 to 902)
de Montigny J., Spehner C., Souciet J., Tekala F., Dujon B., Wincker P., Artiguenave F. and Potier S.

AUTHORS
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila

TITLE
FEBS Lett. 487 (1), 87-90 (2000)

JOURNAL
MEDLINE
20584725

PUBMED
11152890

REFERENCE
3 (bases 1 to 902)
Genoscope.

AUTHORS
Direct Submission

TITLE
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

JOURNAL
This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES
Location/Qualifiers
1..902
/organism="Pichia farinosa"
/mol_type="genomic DNA"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0AA021C12"
/clone_lib="AX0AA"
/note="end : T3"

BASE COUNT
273 a 162 c 166 g 296 t 5 others

ORIGIN

Query Match 25.9%; Score 115.6; DB 29; Length 902;
Best Local Similarity 60.8%; Pred. No. 2.4e-05;
Matches 209; Conservative 1; Mismatches 125; Indels 9; Gaps 1;

QY 13 GATATAGATATGTTTAAATTTAGAGAACCAATATGATTAAGATTTTAAAGAGGT 72
DB 554 GATATAGATATGTTTGGACTTAGAGGATGAGTACTACGATGAGGATACAGGAAGG 613
QY 73 CAATACAGGAACAAAGATCAATATTAGAGAAAGAAATATGGTTATCAAACTGGA 132
DB 614 CAAGGAGTCTTTGTTTCAGGAGCTGCATTTGGAAGGAATGGAATATGTTATCAACAGC 673
QY 133 TTTCAACGATTTTAAATCATTTGGTTTATATTCAGAAATTAATGAATTTGG----- 183
DB 674 ATGCAAGATTTTATAGTTGGTTTACATGAAGGGTTTATGATTACTGGAAGCCAC 733

```

QY 184 TTATCCCATATAGATCAATATAATAACTCTTCTTCACTTCGGATCATTTGTAATTTG 243
D 734 TTGTCTCAGTATGACAAATAGTTAGCATTAAGACACTAGAAATACATTAAGTCTGGCA 793
QY 244 GAAGATATTATGACAAATTTCTATACGAATGGAGATAAAGAAGTTGAAGATTAGAA 303
D 794 GAAGGCAATTTGGATGCGATTTCTATGGAATACGGAAGATCTCTGAAGAGAGTATGAA 853
QY 304 AAAAATATTTAAAGGCAAGAAATAAATTAAGAGTGTATAGTAG 347
D 854 AAAGCCATATAAGACAAATAAATAATTTWGAGTAATAGCAAG 897

RESULT 3
BX414650/c
LOCUS
DEFINITION
  BX414650 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YN02
  3-PRIME, mRNA sequence.
ACCESSION
  BX414650
VERSION
  BX414650.1 GI:30763455
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 994)
  Li W.B., Gruber,C., Jessee,J. and Polayes,D.
  Full-length cDNA libraries and normalization
  Unpublished
  Contact: Genoscope
  Genoscope - Centre National de Sequencage
  BP 191 91006 EVRY cedex - France
  Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
  Library was constructed by Life Technologies, a division of
  Invitrogen. This sequence belongs to sequence cluster 6015.f
  Contact : Feng Liang Email : fliang@lifetech.com URL :
  http://fulllength.invitrogen.com/InvitroGen/Corporation/1600
  Paraday Avenue Genoscope sequence ID : CS0CAP001DG01NP1.
  Location/Qualifiers
  1..994
  /organism="Homo sapiens"
  /mol_type="mRNA"
  /db_xref="taxon:9606"
  /clone="CS0CAP001YN02"
  /tissue_type="THYMUS"
  /clone_lib="Homo sapiens THYMUS"
  /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
  with a NotI-oligo(dT) primer. Five prime end enriched,
  double-strand cDNA was digested with Not I and cloned into
  the Not I and EcoRV sites of the pCMVSPORT 6 vector.
  Library was not normalized."
  BASE COUNT
  185 a 121 c 25 g 324 t 339 others
  ORIGIN
  Query Match
  Best Local Similarity 18.1%; Score 80.8; DB 13; Length 994;
  Matches 77; Conservative 149; Mismatches 108; Indels 0; Gaps 0;

QY 69 AGGTCAATATACAGGACAAAGATCAATATTAGAGGAAAGATATGTTATCAAC 128
D 991 WKWDDDDGRKKDWAWAADKWAARDRWAGARBARADWADWAAARWDRWTAWWA 932
QY 129 TGGATTTCACAGATTTTAAATCATCTGGTTATATTCAGAAATTTTGGTTATC 188
D 931 AKRKKWADWTTKTKKKDADAADGGDDWDATAWMAADWAAARWAAATTTTDDWT 872
QY 189 CCATATAGATCAATATAATACTCTTCTTCACTTCGGATCATTTGGAAGA 248
D 871 AWWTEKDATWKADDADDKATKTTTTTTWKAWTDWAGTAKWADATWADAADAWAA 812
QY 249 TATTATGCGCAAAATTTCTATAACGAATGAGAGATTAAGAGATTTATGAAAAA 308
D 811 RAWAAARAARWAAAWAAADAWWWWWDDWAADAAWAAWAAWTDKDAWRAADWARAAA 752

```

```

QY 309 TATTAAAAAGCAAGAAATAAATTAAGAGTGTATAGTATTAAGAAACCTTGAA 368
D 751 AAWMDWGRGGRADAAGKDKAARAAWWDGAGGRDKDWAARWDDADAAAAAATAATW 692
QY 369 AATTGATTCATTGGATATATTGGTGAAAGAGTA 402
D 691 WATWKKWKDKOWAAAAAAWDTAKGRKDDARDDW 658

```

```

RESULT 4
CNS001FB/c
LOCUS
DEFINITION
  Drosophila melanogaster genome survey sequence TET3 end of BAC #
  BACR04A23 of RPci-98 library from Drosophila melanogaster (fruit
  fly), genomic survey sequence.
ACCESSION
  AL060732
VERSION
  AL060732.1 GI:4939397
KEYWORDS
  GSS.
SOURCE
  Drosophila melanogaster (fruit fly)
ORGANISM
  Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
  Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
  1 (bases 1 to 1101)
  Genoscope.
  Direct Submission
  Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
  BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
  Web : www.genoscope.cns.fr)
  Determination of this BAC-end sequence was carried out as part of a
  collaboration with the Berkeley Drosophila Genome Project (BDGP).
  The BDGP is constructing a physical map of the Drosophila
  melanogaster genome using these BACs. For further information
  please see http://www.fruitfly.org The BDGP Drosophila
  melanogaster BAC library was prepared by Kazuo Ogoe and
  Aaron Mamoer in Pieter de Jong's laboratory in the Department of
  Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
  NY. The library is named RPci-98 and was constructed by partial
  EcoRI digestion of Drosophila DNA and was provided by the BDGP from the
  isogenic strain y2; cn bw sp, the same strain used for the BDGP's
  p1 and EST libraries. A more detailed description of the library
  and how to order individual BAC clones, the entire library, or
  filters for hybridization from the BACPAC Resource Center can be
  found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
  Location/Qualifiers
  1..1101
  /organism="Drosophila melanogaster"
  /mol_type="genomic DNA"
  /db_xref="taxon:7227"
  /clone="BACR04A23"
  /clone_lib="RPci-98"
  /notes="end : TET3"
  BASE COUNT
  288 a 110 c 103 g 491 t 109 others
  ORIGIN

```

```

FEATURES
  source
  Query Match
  Best Local Similarity 17.5%; Score 78.2; DB 29; Length 1101;
  Matches 168; Conservative 48; Mismatches 185; Indels 3; Gaps 1;

QY 4 TCAGATATAGATAGATAGTATGTTTAAATTTAGAGGAAAGAACATATGATTTAGGATTT 63
D 1040 TAAAAAAATATATAATWATTAATAATAATAATWATAAAATWAAAWAAATTAATTAATA 981
QY 64 AAAGAGGTCAATACAGGACAAAGATCAATATTTAGAGGAAAGAAATATGTTAT 123
D 980 TTATWATATTTAAAAAATAAAATTAATAAATAAAATTAATAAATAAATAATAATA 921
QY 124 CAAACTGGATTTCAACGATTTTAAATCATCTGGTTATATTTCAAGAAATTAATGAAATTTGG 183
D 920 TAAAWTAAATWATTAATWATTAATAATAAATAAATAAATAAATAAATAAATAAATA 864
QY 184 TTATCCCATATAGATCAATATAATACTCTTCTTCACTTCGGATCATTTGAAATTTG 243

```


KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 1201)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelosBAC11.

FEATURES
source
1..1201
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACN15M24"
/clone_lib="DrosBAC"
/plasmid="pBelosBAC11"
/notes="end : T7"

BASE COUNT 323 a 87 c 79 g 551 t 161 others

ORIGIN
Query Match 15.9%; Score 71; DB 29; Length 1201;
Best Local Similarity 41.8%; Pred. No. 5.7;
Matches 175; Conservative 42; Mismatches 199; Indels 3; Gaps 1;
QY 6 AGATATAGATAGATAGATGATTAAATTTAGAGAGAGACATATGATAGATTAA 65
DB 1180 AATATATAAATAATATAWAAATATATATWNAANAANTATAAANAATAWATATAA 1121
QY 66 AGAAGGTCAATACAAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGGTTATCA 125
DB 1120 AWAAGAAAWWWATTTAAATAWAAATATAANWAAAAAANAANAANWTTTHTANAA 1061
QY 126 AACTGGATTTCAACGATTTTAAATCATTTGTTATATTCAGAAATTAATGAATTTGGTT 185
DB 1060 TATTWTTTNTATWAAATTTTATTTTATTTTATWAAWAAWAAAAAATAATTTTA 1001
QY 186 ATCCATATAGATCAATATATAACTCTCTCTCACTTCGGAATCATTTGAAATATTTGGA 245
DB 1000 AAAATAAAWTAATATWAAATTTTAAATAATTTTAAATAATTTTAAATAAAWAA 941
QY 246 AGATATTATGCCACAAATTTCTATAACGAATGGAGATAAAGAGTTGAAGATTATGAA-- 303
DB 940 TATWAAATTTTTTTTATWATATAWAAWATTTTTTTTAAATAAAWAAWTTAAAAATW 881
QY 304 -AAAAATATTTAAAGGCAAGAAATAAATTAAGAGTGTATAGTATAGTATACTAAAGAAAC 362
DB 880 TAAAAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 821
QY 363 TTGGAATATGATTCATTTGATTTGTTGGTGAAGAGAGTGGTGAAGATTTCAAGTTA 421
DB 820 TTTWAAAAAANAANAATTTTAAATTTTNTAAATAAANAANAANAANAANAANAATW 762

RESULT 13
CNS00EA9 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR28K21 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL068837
VERSION AL068837.1 GI:4949081

KEYWORDS
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosogawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28K21"
/clone_lib="RPCI-98"
/notes="end : TET3"

BASE COUNT 502 a 8 c 151 g 297 t 143 others

ORIGIN
Query Match 15.8%; Score 70.6; DB 29; Length 1101;
Best Local Similarity 45.0%; Pred. No. 6.6;
Matches 175; Conservative 29; Mismatches 183; Indels 2; Gaps 1;
QY 7 GATATAGATAGATAGATGATTAAATTTAGAGAGAGACATATGATAGATTAA 66
DB 264 GAAAAAATTTGTTGTTAGGAAAGAAATATTAATAAARAAATTTTATAATWATATGATA 323
QY 67 GAAGGTCAATACAAGGAACAAAGATCAATATTTAGAGGAAAGAAATATGGTTATCAA 126
DB 324 GAATATWAAAAAANAATAWATAWAAAAAARAAGAAAGAAAGATTTGATAGAA 383
QY 127 ACTGGATTTCAACGATTTTAAATCATTTGTTATATTCAGAAATTAATGAAATTTGGTTA 186
DB 384 AAGAATATTTGGTGGGTAAAAAANAANAANAANAANAANAANAANAANAANA 443
QY 187 TCCCATATAGATCAATATAATAACTCTCTCTTCCTTCGGAATCATTTGAAATTTGAA 246
DB 444 WATTWAAATTAATWAAATTTAAATAATAWAAATTTTAAANAANAATGATTTTGGTTA 503
QY 247 GATATATGGCACAAATTTCTATAACGAATGGAGATAAAGAGTTGAAGATTATGAAAA 306
DB 504 AATAATATWATWAGTATATGAATWTAATAAANAANAANAANAANAANAANAANA 563
QY 307 AATATTAAGGCAAGAAATAAATAATTAAGAGTGTATAGTATAGTATACTAAAGAACTTGG 366
DB 564 TAAATTAATTTTAAATATGTTTAAAGRTAAATAAGTAAAAAATTTGAAGATATG- 622
QY 367 AAATTTGATTCATTTGGATTAATTTTGGTGAA 395
DB 623 -AAATTTGATTTAAATGATATAAAGGTTTGA 650

RESULT 14
CD049644

LOCUS CD049644 997 bp mRNA linear EST 09-MAY-2003
 DEFINITION AGENCOURT_13986201 NIH_MGC_172 Homo sapiens cDNA 5', mRNA sequence.
 ACCESSION CD049644
 VERSION CD049644.1 GI:30485777
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 997)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: rgs@bbs-rmail.nih.gov
 Tissue Procurement: Dr. Jamie Thompson, University of WI
 cDNA Library Preparation: Gina Zastrow-Hayes
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Cloning Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: NDKM47 row: e column: 22
 High quality sequence start: 10
 High quality sequence stop: 476.
 Location/Qualifiers
 1. .997
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cell"
 /lab_host="DH10B TOR4"
 /clone_lib="NIH_MGC_172"
 /notes="vector: pDONR201; Site 1: attP2; Site 2: attP1;
 LIBR PRIMING - oligo dt; METHOD - full-length enriched;
 Embryonic Stem Cells H1; LIBR PROVIDER - Bradfield"
 427 a 128 c 113 g 306 t 23 others

FEATURES

source
 1. .997
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /tissue_type="embryonic stem cell"
 /lab_host="DH10B TOR4"
 /clone_lib="NIH_MGC_172"
 /notes="vector: pDONR201; Site 1: attP2; Site 2: attP1;
 LIBR PRIMING - oligo dt; METHOD - full-length enriched;
 Embryonic Stem Cells H1; LIBR PROVIDER - Bradfield"
 427 a 128 c 113 g 306 t 23 others

Query Match 15.7%; Score 70.2; DB 14; Length 997;
 Best Local Similarity 46.8%; Pred. No. 7.6;
 Matches 183; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
 QY 6 AGATATAGATATAGATATATGTTTAAATTTAGAGAGAACATATGATTTAGGATTAA 65
 Db AA 452
 QY 66 AGAGGTCAATACAGGACAAAGATCAATTTTAGAGGAGAAAGATATGTTATCA 125
 Db AA 512
 QY 126 AACTGGATTTCAACGATTTTATCAATCGTTTATTTCAAGATTAATGAAATTTTGGTT 185
 Db AA 572
 QY 186 ATCCCATATAGATCAATATTAATCTCTCTTCACCTCGGAATCAATTTGAATTTGA 245
 Db TTATATTAAATTTTAAATAAAAAAAAAATTTTATTTTAAATTTTAAATTTTATTT 632
 QY 246 AGATATTATGCGCAAAATTTCTATAACCAATCGAGATAGAGAGTTGAAGATTATGAAA 305
 Db TTTTAAATATATATATTTTATTAATAATAATAATAATAATAATAATAATAATA 692
 QY 306 AATATTAAAGCGCAAGAAATTAATTAAGAGTGATAGCTAGTATTAATCACTAAAGAACTTG 365
 Db AATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 752
 QY 366 GAAATTTGATTCATTTGGATTAATTTGGTGA 396
 Db TATTATATATTAATTAATAATAATAATAATAATAATAATAATAATAATAATA 783

RESULT 15

Search completed: February 16, 2004, 12:51:31

CNS0155H
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

1. .1001
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone_lib="BACNI3C23"
 /plasmid="pBelobAC11"
 /notes="end : Sp6"

FEATURES

source
 1. .1001
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone_lib="BACNI3C23"
 /plasmid="pBelobAC11"
 /notes="end : Sp6"
 BASE COUNT 266 a 219 c 134 g 150 t 232 others
 ORIGIN

Query Match 15.7%; Score 70; DB 29; Length 1001;
 Best Local Similarity 34.0%; Pred. No. 8.1;
 Matches 134; Conservative 80; Mismatches 180; Indels 0; Gaps 0;
 QY 6 AGATATAGATATAGATATATGTTTAAATTTAGAGAGAACATATGATTTAGGATTAA 65
 Db AAKTKWAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665
 QY 66 AGAGGTCAATACAGGACAAAGATCAATTTTAGAGGAGAAAGATATGTTATCA 125
 Db AARGKAAARRRAAARAAARAGDAARAAARAGRAAARDARAAARAAARAAARAAAR 725
 QY 126 AACTGGATTTCAACGATTTTATCAATCGTTTATTTCAAGATTAATGAAATTTTGGTT 185
 Db AABARDARAAARAAARAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 785
 QY 186 ATCCCATATAGATCAATATTAATCTCTCTTCACCTCGGAATCAATTTGAATTTGA 245
 Db AAWTAAAAAATATTTTAAWAAWAAATTAATAAATTTTAAWAAWAAATTAATAA 845
 QY 246 AGATATTATGCGCAAAATTTCTATACGAATCGAGATAGAGAGTTGAAGATTATGAAA 305
 Db AATTTTAAWAAWAAWAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 905
 QY 306 AATATTAAAGCGCAAGAAATTAATTAAGAGTGATAGCTAGTATTAATCACTAAAGAACTTG 365
 Db WTTTAAWAAWAAWAAATTTTAAWAAWAAATTAATAAATTTTAAWAAWAAATTA 965
 QY 366 GAAATTTGATTCATTTGGATTAATTTGGTGA 399
 Db AAAAAATTAWAAWAAWAAATTAATAAATTAATAAATTAATAAATTAATAA 999

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 16, 2004, 14:27:58 ; Search time 2528 Seconds
(without alignments)
1422.889 Million cell updates/sec

Title: US-09-980-054a-12
Perfect score: 772
Sequence: 1 MSIDIDNVLNLEEEQYELG.....NLVKEVGGLQVSENPDMM 148

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_epool/US0980054/runat_13022004.163840.26309/app_query.fasta_1.327
-DB=EST -QFMT=fastap -SUFFIX=rat -MINMATCH=0.1 -LOOPEXT=0 -COOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cai -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0980054 @CGN 1.1 2810 @runat_13022004.163840.26309 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOC=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:
2: em_esthum:
3: em_estin:
4: em_estnu:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_htc:
9: gb_estl:
10: gb_est2:
11: gb_htc:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: em_gss_hum:
18: em_gss_inv:
19: em_gss_pln:
20: em_gss_vit:
21: em_gss_fun:
22: em_gss_mam:
23: em_gss_mus:
24: em_gss_pro:
25: em_gss_rod:
26: em_gss_phg:
27: em_gss_vrl:
28: gb_gss1:

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	453	58.7	1059	29	CNS07EAO AL441526 T3 end of
c 2	296.5	38.4	902	29	CNS06VIL AL417187 T3 end of
c 3	200.5	26.0	609	29	BZ296387 CG2026.f1
c 4	196.5	25.5	433	29	BZ300548 KD0783.R1
c 5	187	24.2	1030	29	CNS0757A AL431468 T7 end of
c 6	138.5	17.9	849	29	CNS06PHM AL409376 T7 end of
c 7	130	16.8	784	14	CB632715 OS1IEB11C
c 8	122	15.8	606	13	BQ461520 HD04B05r
c 9	122	15.8	611	13	BU968430 HB07H22r
c 10	122	15.8	641	13	BQ472468 HB09H22T
c 11	114	14.8	530	9	AI994649 701499129
c 12	110.5	14.3	879	29	BZ992706 PUBB09TDD
c 13	110	14.2	477	14	CA764105 AF53-Rbf
c 14	110	14.2	566	9	AV913233 AV913233
c 15	109.5	14.2	941	29	CC405283 CB632714 OS1IEB11C
c 16	109.5	14.2	941	29	CC405283 FUHTB67TB
c 17	108	14.0	641	9	AJ432075 AJ432075
c 18	108	14.0	772	10	BF627938 BF627938 HVSMEB000
c 19	107	13.9	630	14	CD431595 ETH1_9.CO
c 20	105.5	13.7	637	10	BF054068 EST439298
c 21	105.5	13.7	634	14	CD038635 UTPPI010
c 22	105	13.6	409	28	AQ962576 LERGG28TR
c 23	105	13.6	571	14	CB885171 3529_1.R2
c 24	104	13.6	661	13	BQ115339 EST600915
c 25	102.5	13.3	533	10	BZ237388 146664.MA
c 26	102.5	13.3	632	10	BG095637 EST460192
c 27	102.5	13.3	645	10	BG599966 EST504861
c 28	102.5	13.3	650	12	BM405065 EST579392
c 29	102.5	13.3	689	13	BQ115338 EST600914
c 30	102.5	13.3	720	12	B432791 EST635552
c 31	101.5	13.1	688	10	BG616931 602615605
c 32	100.5	13.0	366	9	AA304792 EST175780
c 33	100	13.0	569	14	W44346 ZC27B08.r1
c 34	99.5	12.9	458	9	AA148771 Z104F04.r
c 35	99.5	12.9	497	10	BF654530 278745.MA
c 36	99.5	12.9	506	12	BM712319 UI-E-DW1
c 37	99.5	12.9	523	10	BF815233 RCS-C1014
c 38	99.5	12.9	564	13	BQ662888 CL2902.2
c 39	99.5	12.9	565	12	BM761310 K-EST0042
c 40	99.5	12.9	585	13	BX474554 DKEZp686K
c 41	99.5	12.9	589	12	BM849788 K-EST0130
c 42	99.5	12.9	607	10	AW959687 EST371757
c 43	99.5	12.9	687	9	AV682155 AV682155
c 44	99.5	12.9	712	12	BM767075 K-EST0049
c 45	99.5	12.9	716	12	BM850088 K-EST0130

ALIGNMENTS

RESULT 1
CNS07EAO/c
LOCUS CNS07EAO 1059 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone XBD0AA002E07 of library XBD0AA from strain CBS 94
of Candida tropicalis, genomic survey sequence.
ACCSSION AL441526
VERSION AL441526.1 GI:12224752
KEYWORDS GSS.
SOURCE Candida tropicalis
ORGANISM Candida tropicalis
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1059)

AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Boitton-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P., and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 1059)

AUTHORS Blandin, G., Ozier-Kalogeropoulos, O., Wincker, P., Artiguenave, F. and Dujon, B.

TITLE Genomic exploration of the hemiascomycetous yeasts: 16. Candida tropicalis

JOURNAL FEBS Lett. 487 (1), 91-94 (2000)

MEDLINE 20584726

PUBMED 11152891

REFERENCE 3 (bases 1 to 1059)

AUTHORS Direct Submission

TITLE Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: segref@genoscope.cns.fr Web: www.genoscope.cns.fr)

JOURNAL This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES Location/Qualifiers

1..1059

organism="Candida tropicalis"

mol_type="genomic DNA"

strain="CBS 94"

db_xref="taxon:5482"

clone="XBD0AA02E0"

clone_lib="XBD0AA"

note="end: T3"

misc_feature complement (<26..>346)

notes="similar to Saccharomyces cerevisiae ORF YNL260c [hypothetical protein]"

evidence="not experimental"

BASE COUNT 362 a 149 c 198 g 348 t

ORIGIN

Alignment Scores:

Pred. No.: 7,12e-45 Length: 1059

Score: 453.00 Matches: 85

Percent Similarity: 81.15% Conservative: 14

Best Local Similarity: 69.67% Mismatches: 23

Query Match: 58.68% Indels: 0

DB: 29 Gaps: 0

US-09-980-054a-12 (1-148) x CNS07EAO (1-1059)

Qy 1 MetSerAspIleAspIleValLeuAsnLeuGluGluGlnTyrGluLeuGly 20

Db 367 ATGAGCGATTTTACATATATGAAGTATTAACCTAGAGAGAACATATAATTTAGGA 308

Qy 21 PheLysGluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyGlnTyrGly 40

Db 307 TTTAAGAGGGACAGGAGCACTCTACCAAGAACATATCTTAGAGGAAACAATATGTT 248

Qy 41 TyrGlnThrGlyPheGlnArgPheLeuIleIleGlyTyrIleGlnIleuLeuMetLysPhe 60

Db 247 TATCAAACTGGATTCGAAGGTTTAAATGTTGGCTATATTCAGGGATTTGGTAGCGAA 188

Qy 61 TrpLeuSerHisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsn 80

Db 187 TGGTTGGATAAATCTAGACCAACTATAATGTTCCAAATCACTTCAAGGTCATATAATCAA 128

Qy 81 LeuGluAspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyr 100

Db 127 CTAAAGTGAATATACTGACATACCATTCGACAAATGGAGATGAAGAGTTGAAAATAC 68

Qy 101 GluLysAsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGlu 120

Db 67 GAAAAAACAATCAAAAGGCAAGAAATAAATAAGAGTGAATTCGAACATACTAAAGAA 8

Qy 121 ThrTyr 122

Db 7 AATGG 2

RESULT 2

CNS06VIL

LOCUS T3 end of clone AX0AA021C12 of library AX0AA from strain CBS 7064

DEFINITION of Pichia farinosa, genomic survey sequence.

ACCESSION AL417187

VERSION AL417187.1 GI:12198580

KEYWORDS GSS.

SOURCE Pichia farinosa

ORGANISM Pichia farinosa

REFERENCE 1 (bases 1 to 902)

AUTHORS Souciet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Boitton-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuveglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.

TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

JOURNAL FEBS Lett. 487 (1), 3-12 (2000)

MEDLINE 20584711

PUBMED 11152876

REFERENCE 2 (bases 1 to 902)

AUTHORS de Montigny, J., Spehner, C., Souciet, J., Tekala, F., Dujon, B., Wincker, P., Artiguenave, F. and Potier, S.

TITLE Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila

JOURNAL FEBS Lett. 487 (1), 87-90 (2000)

MEDLINE 20584725

PUBMED 11152890

REFERENCE 3 (bases 1 to 902)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: segref@genoscope.cns.fr Web: www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES Location/Qualifiers

1..902

organism="Pichia farinosa"

mol_type="genomic DNA"

strain="CBS 7064"

db_xref="taxon:4920"

clone="AX0AA021C12"

clone_lib="AX0AA"

note="end: T3"

BASE COUNT 273 a 162 c 166 g 296 t 5 others
ORIGIN

Alignment Scores:

Pred. No.: 8,28e-26 Length: 902
Score: 296.50 Matches: 58
Percent Similarity: 71.55% Conservative: 25
Best Local Similarity: 50.00% Mismatches: 30
Query Match: 38.41% Indels: 3
DB: 29 Gaps: 1

US-09-980-054A-12 (1-148) x CNS06VIL (1-902)

QY 5 AspileAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGluGly 24
DB 554 GATATAGAGATGTTGGACTTATAGAGATGAGTACTACGATGAGGATACAGGAGG 613
QY 25 GlnileGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGly 44
DB 614 CAAGGAGTCTTTGTTTCAGGAGCTCATTTGAGGAGATGGAATATGTTTATCAACAGGC 673
QY 45 PheGlnArgPheLeuileileGlyTyrileGlnGluLeuMetLysPheTyrPleuSerHis 64
DB 674 ATGCAAGATTTTAAATAGTGGTTATGATGAAGCGTTAATGATTTACTGGAAGACCC 733
QY 65 IleAspGlnTyrAsn-----AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeu 81
DB 734 TTGTCAGTATGAGCAATATAGTATGAGCACTAGAAAATCATTAACTCTGGCA 793
QY 82 GluAspMetAlaGlnileSerileThrAsnGlyAspLysGluValGluAspTyrGlu 101
DB 794 GAAGCATTTGGATGGCATTTCTATGGANATACGGAATGCTGTAAAGAGATATGAA 853
QY 102 LysAnileLysLysAlaArgAsnLysLeuArgValleAlaSerille 117
DB 854 AAAGCCATCACTAAAGAGCAAAATAAATTTGAGTAAATAGCAAGCTCT 901

RESULT 3

BZ296387/C 509 bp DNA linear GSS 31-OCT-2002
LOCUS
DEFINITION CG2026.f11 Candida glabrata Random Genomic Library Candida glabrata genomic clone CG2026, genomic survey sequence.

BZ296387

VERSION BZ296387.1 GI:24438257

KEYWORDS GSS.

SOURCE Candida glabrata

ORGANISM Candida glabrata

REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; mitosporic Saccharomycetales; Candida.

1 (bases 1 to 609)

Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

Evidence from comparative genomics for a complete sexual cycle in

the 'asexual' pathogenic yeast Candida glabrata

Genome Biol. 4 (2), R10 (2003)

JOURNAL

MEDLINE

PUBMED

12620120

CONTACT: Wong S

Department of Genetics, Smurfit Institute

Dublin 2, Ireland

Tel: 353 1 6082319

Fax: 353 1 6798558

Email: swong@cd.ie

Class: plasmid ends.

Location/Qualifiers

1..609

/organism="Candida glabrata"

/mol_type="genomic DNA"

/strain="CBS 138"

/db_xref="taxon:5478"

/clone="CG2026"

/clone_lib="Candida glabrata Random Genomic Library"

183 a 127 c 85 g 214 t

BASE COUNT

ORIGIN

Alignment Scores:

Pred. No.: 2,78e-14 Length: 609
Score: 200.50 Matches: 43
Percent Similarity: 57.64% Conservative: 40
Best Local Similarity: 29.86% Mismatches: 50
Query Match: 25.97% Indels: 11
DB: 29 Gaps: 3

US-09-980-054A-12 (1-148) x BZ296387 (1-609)

QY 4 IleAspileAspAsnValLeuAsnLeuGluGluGluGlnTyrGluLeuGlyPheLysGlu 23
DB 591 ATGATCTAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 532
QY 24 GlyGlnileGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThr 43
DB 531 GCGAGAAATGAATAACTTGAAGAACATTTAATAGAGGTAAGAGTTTGGTTTACAAGTT 472
QY 44 GlyPheGlnArgPheLeuileileGlyTyrileGlnGluLeuMetLysPheTyrPleuSer 63
DB 471 CGATTTCAAGATTTTGGTTTACTAGGTCAAATGGTTGGGATGTGTATGTC---CTTGAT 415
QY 64 HisIleAspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
DB 414 TCTTTACAGTTTGAAGATAAT-----TCGTTATTTAAGAACATCGCAATGTTCAGAGAG 361
QY 84 IleMetAlaGlnileSerileThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
DB 360 CTGATCAGCTCTATCGAGCTTAACACAAAGAGAGAGATGTTGAAGATTAGAAAAGACT 301
QY 104 IleLysLysAlaArgAsnLysLeuArgValleAlaSerileThrLysGluThrLys 123
DB 300 CTAGTGAATTTAAAGATAAATTTAGATTGATTTTCTTTCATTCAGAGGGAATTTAAA 241
QY 124 -----IleAspSerLeuAsnLeuValLysGluValGly 135
DB 240 CACACAGATCGCAGCTATCAATTTGATTAATTTGAGAGACTTGTCAAGAGCAATTGCA 181
QY 136 GlyThrLeuGln 139
DB 180 GGAGAAATCAAA 169

RESULT 4

BZ300548

LOCUS

DEFINITION

433 bp DNA linear GSS 31-OCT-2002

KB0783.R1 Kluyveromyces delphensis

sequence.

BZ300548

VERSION

BZ300548.1

KEYWORDS

GSS.

SOURCE

ORGANISM

Kluyveromyces delphensis

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

1 (bases 1 to 433)

Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.

Evidence from comparative genomics for a complete sexual cycle in

the 'asexual' pathogenic yeast Candida glabrata

Genome Biol. 4 (2), R10 (2003)

JOURNAL

MEDLINE

PUBMED

12620120

CONTACT: Wong S

Department of Genetics, Smurfit Institute

Trinity College Dublin

Dublin 2, Ireland

Tel: 353 1 6082319

Fax: 353 1 6798558

Email: swong@cd.ie

Class: plasmid ends.

Location/Qualifiers

FEATURES

```

source
1. .433
/organism="Kluyveromyces delphensis"
/mol_type="genomic DNA"
/strain="CBS 2170"
/db_xref="taxon:51657"
/clone_lib="Kluyveromyces delphensis Random Genomic Library"
BASE COUNT 161 a 54 c 87 g 131 t
ORIGIN

Alignment Scores:
Pred. No.: 5,47e-14 Length: 433
Score: 196.50 Matches: 41
Percent Similarity: 60.33% Conservative: 32
Best Local Similarity: 33.88% Mismatches: 43
Query Match: 25,45% Indels: 5
DB: 29 Gaps: 2

US-09-980-054A-12 (1-148) x B2300548 (1-433)
QY 4 ileAspIleAspAsnValleuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGlu 23
Db 49 ATGGATGTGAGCTGTTGTTTATTTAGAGGAGTTTTACAAAGAGGCTCAAGAA 108
QY 24 GlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThr 43
Db 109 GGTGAAGTGAATAATTAAGAACAACTCTCTGAAGGTAAGGAATTTGGTTTCAAGTA 168
QY 44 GlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTyrLeuSer 63
Db 169 GGTTCACAGCTTCTGCTCTCTGCTCAANGTTAGGACTATGTGTGTTGTTGTTG 219
QY 64 HisIleAspGlnTyrAsn---AsnSerSerLeuArgAsnHisLeuAsnAsnLeuGlu 82
Db 220 ---CTGGATTCCTTCACTTGAAGATAAATCACTCGCAAAAACGTCGCAATAATAAGA 276
QY 83 AspIleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLys 102
Db 277 GAACATACACACATACATTCATACACAGGAGGAATGTGGAGGCGCTTGAAGAA 336
QY 103 AsnIleLysLysAlaArgAsnLysLeuArgValIleAlaSerIleThrLysGluThrTyr 122
Db 337 AAATGGTTAAACTTAAATAAAGATATAGTATATATTGCTAGCTTTTCAAGGGATAT 396
QY 123 Lys 123
Db 397 AAA 399

RESULT 5
CNS076JA/c 1030 bp DNA linear GSS 07-JUL-2001
LOCUS T7 end of clone BB0AA004B04 of library BB0AA from strain CBS 4732
DEFINITION of Pichia angusta, genomic survey sequence.
ACCESSION AL431468
VERSION AL431468.1 GI:12214880
KEYWORDS GSS.
SOURCE Pichia angusta
ORGANISM Pichia angusta

REFERENCE
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Maupertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekalia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
Yeast species for molecular evolution studies
PDBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876

REFERENCE
AUTHORS Blandin,G., Llorente,B., Maupertuy,A., Wincker,P., Artiguenave,F.
and Dujon,B.
Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
PDBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
PUBMED 11152888
REFERENCE
AUTHORS 3 (bases 1 to 1030)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail:
secref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
Location/Qualifiers
1. .1030
/organism="Pichia angusta"
/mol_type="genomic DNA"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone_lib="BB0AA004B04"
/clone_lib="BB0AA"
/note="end: T7"
misc_feature complement(<43..>222)
/note="similar to Saccharomyces cerevisiae ORF YNL260c [
hypothetical protein]"
/evidence="not-experimental"
BASE COUNT 318 a 192 c 302 g 216 t
ORIGIN

Alignment Scores:
Pred. No.: 2,51e-12 Length: 1030
Score: 187.00 Matches: 32
Percent Similarity: 76.67% Conservative: 14
Best Local Similarity: 53.33% Mismatches: 14
Query Match: 24,22% Indels: 0
DB: 29 Gaps: 0

US-09-980-054A-12 (1-148) x CNS076JA (1-1030)
QY 3 AspIleAspIleAspAsnValleuAsnLeuGluGluGlnTyrGluLeuGlyPheLys 22
Db 216 GACATAGACTTCGATTCGCTGTGAACCTTGAGACCACTACTATCAGGAAGGTTTCATC 157
QY 23 GluGlyGlnIleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyThr 42
Db 156 GAAGGCGACGTCGAGGGCTCGAACAACAGTTTTTAGAGGAAGCAACTTGGCATCCAG 97
QY 43 ThrGlyPheGlnArgPheLeuIleGlyTyrIleGlnGluLeuMetLysPheTyrLeu 62
Db 96 ACCGGTTTTCCAGCGTCTTCTGTCCTCGTCAAGTATAGTATAGTATAGTATAGTAT 37

RESULT 6
CNS06PHM 849 bp DNA linear GSS 05-JUL-2001
LOCUS T7 end of clone AV0AA014E08 of library AV0AA from strain CBS 379 of
DEFINITION Saccharomyces exiguus, genomic survey sequence.
ACCESSION AL409376
VERSION AL409376.1 GI:12176838
KEYWORDS GSS.
SOURCE Saccharomyces exiguus
ORGANISM Saccharomyces exiguus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```



```

Db      593 TTATGATGACACACATATCAAAAGGGTTTCAAGATGGTTATAGTGAAGGCTTGGTGTCT 534
QY      32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 51
Db      533 GGAAAGAGAGAGGAGGAGGTTGGTTTAAAGATGGTTTTCAGGTAGGTGAAGAAGCTA 474
QY      52 GlyTyrIleGlnGluLeuMetLysPheTyrLeuSer-----HisIleAspGlnTyrAsn 69
Db      473 GGTGTTTATCAGGGATGCTCGATGTTTGGACGTGTTGGTGGTTTCAATTGATCAAGATGCA 414
QY      70 AsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAspIleMetAlaGlnIleSer 89
Db      413 TTTCTCAGCTCGGTCAGGAAACACATGAGCAACTAGCTGCACCTCTTGAGAGCTATCCG 354
QY      90 IleThrAsn---GlyAspLysGluValGluAspTyrGlyLysAsnIleLysLysAlaArg 108
Db      353 TTGTGCAACCCAGAGACGAGCAAGTTCAGATATA-----ATGGAGAGATAGA 303
QY      109 AsnLysLeuArgValIleAlaSer-IleThrLysGluThrTrpLysIle 124
Db      302 CTGAATTCAGGGTTATCACAGCAAGTTTAGGTACAAAACCTGGAGTATC 254

RESULT 8
BQ461520
LOCUS   BQ461520
DEFINITION HD04B05: HD Hordeum vulgare cDNA clone HD04B05 5-PRIME, mRNA
SEQUENCE
ACCESSION BQ461520.1 GI:21269291
VERSION   BQ461520.1
KEYWORDS  EST.
SOURCE    Hordeum vulgare
ORGANISM  Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
Zhang,H., Weschke,W., Michalek,W., Stein,N. and Graner,A.
EST sequencing and analysis in barley (2002)
Unpublished
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 606 Std Error: 0.00
Plate: 4 row: B column: 5
Seq primer: M13rev.
Location/Qualifiers
1..606
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Golden Promise"
/db_xref="taxon:4513"
/clone="HD04B05"
/tissue_type="callus"
/dev_stage="callus (5-10 mm in diameter)"
/lab_host="XL10-Gold"
/clone_lib="HD"
/note="Vector: Bluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning
artifact caused by the kit, in most cases the EcoRI site
is NOT present, as well as the EcoRI adapter used for
cloning. To excise the insert, restriction sites upstream
EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also
due to the cloning system used Blue/white selection for
recombinants is not 100% reliable.Average insert size is 1
kb"
172 a 118 c 174 g 142 t

```

```

Alignment Scores:
Pred. No.: 0.000113 Length: 606
Score: 122.00 Matches: 32
Percent Similarity: 51.35% Conservative: 25
Best Local Similarity: 28.83% Mismatches: 50
Query Match: 15.80% Indels: 4
DB: 13 Gaps: 2

US-09-980-054A-12 (1-148) x BQ461520 (1-606)

QY      6 IleAspAsnValLeuAsnLeuGluGluGlnTyrGluLeuGlyPheLysGluGln 25
Db      258 CTTGAACCAATGTTAGCTTTAGACGAGACACATTATCAAGATGTTTACAAGATGTTAT 317
QY      26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
Db      318 GATGATGGCTTGTATCTGGGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 377
QY      46 GlnArgPheLeuIleIleGlyTyrIleGlnGluLeuMetLysPheTyrLeuSerHisIle 65
Db      378 CAGGTAGGTGAGAGCTAGGATTTCTATCAGGGCTGCTTGGATGTGTGGATGTCAATAATT 437
QY      66 -----AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
Db      438 TGCCTTGTATCAAGATGATTTCTCAGCTCGGTCAGGAAACACATGAGCAATTAGCTGCA 497
QY      84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGlyLysAsn 103
Db      498 CTTCTAAGCAACTATCTTTGTCT-----GATCCAGAGATAATCAGCTTCAAGACATG 551
QY      104 IleLysLysAlaArgAsnLysLeuArgValIle 114
Db      552 ATGAAGGATATAGGCTGAATTCAGGTTATC 584

RESULT 9
BQ968430
LOCUS   BQ968430
DEFINITION HB07H22r BC Hordeum vulgare subsp. vulgare cDNA clone HB07H22
5-PRIME, mRNA sequence.
ACCESSION BQ968430
VERSION   BQ968430.1 GI:24219223
KEYWORDS  EST.
SOURCE    Hordeum vulgare subsp. vulgare
ORGANISM  Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 611)
Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
Barley ESTs from developing seeds
Unpublished
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 611 Std Error: 0.00
Plate: 7 row: H column: 22
Seq primer: M13rev.
Location/Qualifiers
1..611
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/db_xref="GABI:233254"
/db_xref="taxon:112509"
/clone="HB07H22"
/tissue_type="developing caryopsis"
/dev_stage="8-15 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="BC"

```

FEATURES
source

BASE COUNT
ORIGIN

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); developing caryopsis , 8-15 DAP(days after pollination) Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

BASE COUNT 171 a 119 c 178 g 143 t
ORIGIN

Alignment Scores:
Pred. No.: 0.000114 Length: 611
Score: 122.00 Matches: 32
Percent Similarity: 51.35% Conservative: 25
Best Local Similarity: 28.83% Mismatches: 50
Query Match: 15.80% Indels: 4
DB: 13 Gaps: 2

US-09-980-054A-12 (1-148) x BU968430 (1-611)

Qy 6 IleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
Db 260 CTTCGACCAATGGTAGCTTAGACGAGACACATTATCAGATGGTTACAAAGATGTTAT 319
Qy 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
Db 320 GATGATGGCTTGGTATCTGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 379
Qy 46 GlnArgPheLeuLeuIleGlyTyrIleGlnGluLeuMetLysPheLysSerHisle 65
Db 380 CAGTAGGTGAGAGCTAGATGTTCTATCAGGGCTGCTGATGTTGGATGTCATATAT 439
Qy 66 -----AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
Db 440 TGCCTTGCATCAGATGATCTCAGCTCGGGTCAGGAAACACATAGGCAATTAGCTGCA 499
Qy 84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
Db 500 CTCTTAAGCAACTATCCTTTGTCT-----GATCCAGAGAATAATCAGCTTCAAGCATG 553
Qy 104 IleLysLysAlaArgAsnLysLeuArgValle 114
Db 554 ATGAAGGATATAAGCTGAATTCAGGGTTATC 586

RESULT 10
BQ472468
LOCUS BQ472468 641 bp mRNA linear EST 31-MAY-2002
DEFINITION HB09H22T BC Hordeum vulgare subsp. vulgare cDNA clone HB09H22
5-PRIME, mRNA sequence.

ACCESSION BQ472468
VERSION BQ472468.1 GI:21284597
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.

1 (bases 1 to 641)
Rachuk, V., Zhang, H., Weschke, W., Potokina, E. and Wobus, U.
Barley ESTs from developing seeds
Unpublished

Contact: Stejneger Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595

Email: stejnepk-gatersleben.de
Insert Length: 641 Std Error: 0.00
Plate: 9 row: H column: 22
Seq primer: T3.

FEATURES

source

Location/Qualifiers
1..641
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivat="barke"
/db_xref="taxon:112509"
/clone="HB09H22"
/tissue_type="developing caryopsis"
/dev_stage="8-15 DAP (days after pollination)"
/lab_host="XL10-Gold"
/clone_lib="BC"
/notes="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); developing caryopsis , 8-15 DAP(days after pollination) Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

BASE COUNT 177 a 127 c 183 g 154 t
ORIGIN

Alignment Scores:
Pred. No.: 0.000121 Length: 641
Score: 122.00 Matches: 32
Percent Similarity: 51.35% Conservative: 25
Best Local Similarity: 28.83% Mismatches: 50
Query Match: 15.80% Indels: 4
DB: 13 Gaps: 2

US-09-980-054A-12 (1-148) x BQ472468 (1-641)

Qy 6 IleAspAsnValLeuAsnLeuGluGlnTyrGluLeuGlyPheLysGluGlyGln 25
Db 260 CTTCGACCAATGGTAGCTTAGACGAGACACATTATCAGATGGTTACAAAGATGTTAT 319
Qy 26 IleGlnGlyThrLysAspGlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPhe 45
Db 320 GATGATGGCTTGGTATCTGGAGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 379
Qy 46 GlnArgPheLeuLeuIleGlyTyrIleGlnGluLeuMetLysPheLysSerHisle 65
Db 380 CAGTAGGTGAGAGCTAGATGTTCTATCAGGGCTGCTGATGTTGGATGTCATATAT 439
Qy 66 -----AspGlnTyrAsnAsnSerSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsp 83
Db 440 TGCCTTGCATCAGATGATCTCAGCTCGGGTCAGGAAACACATAGGCAATTAGCTGCA 499
Qy 84 IleMetAlaGlnIleSerIleThrAsnGlyAspLysGluValGluAspTyrGluLysAsn 103
Db 500 CTCTTAAGCAACTATCCTTTGTCT-----GATCCAGAGAATAATCAGCTTCAAGCATG 553
Qy 104 IleLysLysAlaArgAsnLysLeuArgValle 114
Db 554 ATGAAGGATATAAGCTGAATTCAGGGTTATC 586

RESULT 11
AI994649
LOCUS AI994649 530 bp mRNA linear EST 08-SEP-1999
DEFINITION 701499129 A. thaliana, Ohio State clone set Arabidopsis thaliana
cDNA clone 701499129, mRNA sequence.

ACCESSION AI994649
VERSION AI994649.1 GI:5841554
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids
; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 530)
Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D.,
Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P.,


```

LOCUS       CB632714               648 bp     mRNA     linear     EST 08-APR-2003
DEFINITION  OSIIEB11C06.f OSIIEB Oryza sativa (indica cultivar-group) cDNA
            clone OSIIEB11C06 5', mRNA sequence.
ACCESSION   CB632714
VERSION     CB632714.1  GI:29627703
KEYWORDS    EST.
SOURCE      Oryza sativa (indica cultivar-group)
            Oryza sativa (indica cultivar-group)
ORGANISM    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzoideae; Oryza.
REFERENCE   1 (bases 1 to 648)
AUTHORS    Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
            Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE       Large-scale identification of ESTs involved in the interaction
            between rice and Magnaporthe grisea
JOURNAL     Unpublished
COMMENT     Contract: Rod Wing
            Arizona Genomics Institute
            University of Arizona
            Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
            85721-0088, USA
            Tel: 520 626 3967
            Fax: 520 621 9288
            Email: http://genome.arizona.edu
PCR PRIMERS
FORWARD: gta aaa cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 11 row: C column: 06
Seq primer: gta aaa cga cgg cca gtc.
FEATURES             Location/Qualifiers
     source           1..648
                     /organism="Oryza sativa (indica cultivar-group)"
                     /mol_type="mRNA"
                     /cultivar="IR36"
                     /db_xref="taxon:39946"
                     /clone="OSIIEB11C06"
                     /tissue_type="Leaf"
                     /dev_stage="3 week"
                     /lab_host="DH10B"
                     /clone_lib="OSIIEB"
                     /note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:
                     XhoI; 24 hrs after inoculation with Rice Blast (PO6-6-3)"
BASE COUNT   155 a   150 c   182 g   161 t
ORIGIN
Alignment Scores:
Pred. No.:      0.00362      Length:      648
Score:          110.00      Matches:    26
Percent Similarity: 50.57%      Conservative: 18
Best Local Similarity: 29.89%      Mismatches:  41
Query Match:    14.25%      Indels:      2
DB:             14      Gaps:        1

US-09-980-054A-12 (1-148) x CB632714 (1-648)
Qy      12 LeuGluGluGlnTyrGluLeuGlyPheLysGluGlyGlnIleGlnGlyThrLysAsp 31
Db      378 TTAGATGAGACACACTATCAACGGTTTCAGAAATGCTTATAGTGAAGGCTTGGTGCT 437
Qy      32 GlnTyrLeuGluGlyLysGluTyrGlyTyrGlnThrGlyPheGlnArgPheLeuIle 51
Db      438 GGAAGAAGAGAGGGAAGGAGGTTGGTTAAAGAAATGCTTTTCAGGTAGGTGAAGAACTA 497
Qy      52 GlyTyrIleGlnGluLeuMetLysPheTyrLeuSer-----HisIleAspGlnTyrAsn 69
Db      498 GGTTTTTATCAGGATGCTCGGATGTTTGGACGTCGTTGGTTTCATTGATCAAGATGCA 557
Qy      70 AsnSerSerLeuArgAsnHisLeuAsnAsnLeuGluAsnMetAlaGlnIleSer 89
Db      558 TTCTCAGCTCGGTGAGGAAAAAATTCAGCAACTAGTCTGACTGTTGAGAGCTATCCG 617
Qy      90 IleThrAsnGlyAspLysGlu 96

```

Db 618 TTGTCGACCCAGAGACGAG 638

Search completed: February 16, 2004, 16:07:51
Job time : 2539 secs

Result	Query	Score	Match	Length	DB	ID	Description
1		132	17.1	134	3	Q97Q1	Q97q1 schizosacch
2		129	16.7	144	10	Q8GV9	Q8GV9 cryza sativ
3		115.5	15.0	139	5	Q81x1	Q81x1 dictyosteli
4		99.5	12.9	226	4	Q8WV25	Q8WV25 homo sapien
5		99.5	12.9	226	4	Q9NR11	Q9NR11 homo sapien
6		99	12.8	222	2	Q9SOB1	Q9SOB1 borrelia bu
7		96	12.4	179	2	Q9SO39	Q9SO39 borrelia bu
8		96	12.4	223	16	Q8DJ3E3	Q8DJ3e3 wigglewort
9		95.5	12.4	141	5	Q9V670	Q9V670 drosophila
10		95	12.3	1680	4	Q9P129	Q9P129 homo sapien
11		94.5	12.2	192	2	Q8S082	Q8S082 borrelia bu
12		92	11.9	270	16	Q9K070	Q9K070 vibrio chol
13		91.5	11.9	5251	5	Q811D4	Q811d4 plasmodium
14		91	11.8	239	16	Q8Z106	Q8Z106 versinia pe
15		91	11.8	247	16	Q8CKH8	Q8CKH8 versinia pe
16		91	11.8	266	2	Q9Z6G6	Q9Z6G6 vibrio para


```

RESULT 6
Q9S0B1 PRELIMINARY; PRT; 222 AA.
AC Q9S0B1
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Conserved hypothetical protein.
GN BBN34.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Plasmid cp32-6.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
SEQUENCE FROM N.A.
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,
RA White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van-Vugt R.,
RA Palmer N., Haft D., Rosa P., Stevenson B.;
RA "A bacterial genome in flux: The twelve linear and nine circular
RA extrachromosomal DNAs in an infectious isolate of the Lyme disease
RA spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 0:0-0(1999).
DR EMBL; AF001578; AAF07545.1; -.
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF02524; KID; 9.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 222 AA; 25444 MW; 1E7F5A0EAA303BDE CRC64;

Query Match 12.8%; Score 99; DB 2; Length 222;
Best Local Similarity 30.2%; Pred. No. 2.8;
Matches 32; Conservative 22; Mismatches 48; Indels 4; Gaps 4;

QY 37 KEYGVTGFORFLIIGYIOELMKFWLSHIDQYNNSSSLRNHLNLEDIMAOISITNGDKE 96
Db 7 KTYNIESIKNEFLNIGFSKDAIDFVFLHNDY-NFEFLKEIKIDIEKND-QKDISNLDK 64
QY 97 VEDYKNIKKARNLRV-IASITKE-TWKDSDNLNVEKVGTTQV 140
Db 65 IDNVEKNLNKIDGNLKNIDNVEKSLNAKIDSLDTKIDNVEKNLNI 110

RESULT 7
Q9S039 PRELIMINARY; PRT; 179 AA.
AC Q9S039
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Conserved hypothetical protein.
GN BBN34.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Plasmid cp32-9.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
SEQUENCE FROM N.A.
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Lathigra R.,
RA White O., Dodson R., Hickey E.K., Gwinn M., Peterson J., van-Vugt R.,
RA Palmer N., Haft D., Rosa P., Stevenson B.;
RA "A bacterial genome in flux: The twelve linear and nine circular
RA extrachromosomal DNAs in an infectious isolate of the Lyme disease
RA spirochete Borrelia burgdorferi.";
RL Mol. Microbiol. 0:0-0(1999).
DR EMBL; AF001581; AAF07675.1; -.
DR InterPro; IPR003900; KID_repeat.
DR Pfam; PF02524; KID; 7.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 179 AA; 20713 MW; 40C1D64732A5CC8A CRC64;

Query Match 12.4%; Score 96; DB 2; Length 179;
Best Local Similarity 31.2%; Pred. No. 3.8;

```

```

Matches 34; Conservative 20; Mismatches 47; Indels 8; Gaps 5;

QY 37 KEYGVTGFORFLIIGYIOELMKFWLSHIDQYNNSSSLRNHLNLEDIMAOISITNGDKE 96
Db 7 RTYDIESIKNEFLNIGFSKDAIDFVFLHNDY-NFEVLKEIKIDVEKNL-QKDISSLDTK 64
QY 97 VEDYKNIKKARNLRV-IASITKE-TWKDSDNLNVEKVGTTQV 139
Db 65 IDNVEKNLNKIDGNLKNIDNVEKSLNAKIDSLDTKIDNVEKNLNI 113

RESULT 8
Q9D3E3 PRELIMINARY; PRT; 223 AA.
AC Q9D3E3
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE FLIH protein.
GN FLIH.
OS Wigglesworthia brevipalpis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Wigglesworthia.
OX NCBI_TaxID=164609;
RN [1]
SEQUENCE FROM N.A.
RA MEDLINE=2297718; PubMed=12219091;
RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
RA Aksoy S.;
RA "Genome sequence of the endocellular obligate symbiont of tsetse
RA flies, Wigglesworthia glossinidia.";
RL Nat. Genet. 32:402-407(2002).
DR EMBL; AB063521; BAC24204.1; -.
KW Complete proteome.
SQ SEQUENCE 223 AA; 25650 MW; CC5697B9884937B0 CRC64;

Query Match 12.4%; Score 96; DB 16; Length 223;
Best Local Similarity 27.6%; Pred. No. 4.9;
Matches 40; Conservative 27; Mismatches 46; Indels 32; Gaps 7;

QY 7 DNVLNLEEBQYELGFKEGQIQGTQDQYLGKGYGVTGQFQFLIIGYIOELMKFWLSHID 66
Db 48 DENLNYE---KIGFESG-----YKKGKIGYKKGKGLI--EIKNEKKKKILQIE 93
QY 67 QYNNSSSLRNHLNLEDI---MAQISITNGKVEDYKNIKKARNLRVIASITKTETW 122
Db 94 KP-LVELKSLIGLDSVIPSKIMQIAIKISEKIK-----QSPLCNTKVLKETO 142
QY 123 KIDSLNVLKVGTTQVSENPDMM 147
Db 143 KLLHNKNF---LGNLSIYVHPDDI 164

RESULT 9
Q9V670 PRELIMINARY; PRT; 141 AA.
AC Q9V670
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE CG13175 protein.
GN CG13175 OR CG13174.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
SEQUENCE FROM N.A.
RA STRAIN=Berkely;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,

```

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Arif J.F., Agbayani A., An H.-J., Andrews-fankoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Benson K.F., Beran P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Berkova D., Borchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.R., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Fesler C., Gabriellian A.B., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.B.,
 RA Nelson D.R., Nelson K.A., Nixon K., Ruskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rainert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.F., Zaveri J.S., Zhan M., Zhou X., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cealnikar S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Ananides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dorsett K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunco J.,
 RA Pacleb J., Fargas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RA "Sequencing of *Drosophila melanogaster* genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminer J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Beran B., Carlson J.W., Cealnikar S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of *Drosophila melanogaster* genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Cealnikar S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE003823; AAF58560.2; -

DR FlyBase; FBgn0033693; CG131175.
 SQ SEQUENCE 141 AA; 15724 MW; CA0016D1658BAC84 CRC64;
 Query Match 12.4%; Score 95.5; DB 5; Length 141;
 Best Local Similarity 25.5%; Pred. No. 3.2; Indels 11; Gaps 5;
 Matches 28; Conservative 29; Mismatches 42;
 QY 7 DNVNLBEQYELGPKQIQGTQDYVLEKGYGYQTFORFLIYQIQLMKFWLSH- 65
 DB 13 DDIVLTKEARLGYEE---GLKQGGQNGNEGYKLGVAQVSLG--BELGKI-LGQVV 65
 QY 66 --DOYNNSSSLRNHLNLEDDIMAQISITNGDK-EVEDYEKNIKKARKLR 112
 DB 66 AQQLKHTDKVRSLEQLRSLEPRTPNDPQADIVGAVQIRSHRRLR 115
 ID Q9P129 PRELIMINARY; PRT; 1680 AA.
 AC Q9P129;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Hypothetical protein KIAA1529 (Fragment).
 GN KIAA1529.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 ON NCBI_TaxID=9606;
 RN [1]_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20277482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human
 RT genes.XVII. The complete sequences of 100 new cDNA clones from brain
 RT which code for large proteins in vitro.";
 RL DNA Res. 7:143-150 (2000).
 DR EMBL; AB040962; BAA96053.1; -
 KW Hypothetical protein.
 FT NON-TER
 SQ SEQUENCE 1680 AA; 194698 MW; BE3AF4E4E0013288 CRC64;
 Query Match 12.3%; Score 95; DB 4; Length 1680;
 Best Local Similarity 25.2%; Pred. No. 60;
 Matches 36; Conservative 27; Mismatches 40; Indels 40; Gaps 7;
 QY 24 GQIQGTQDYVLEKGYGY-----QTGQRFLLIYQIQLMKFWLSH----- 64
 DB 660 GALQKVBEDLELDKSPETLADQTEWQSSLELFYFOEVVQVLMFAHQSELLVQLELEKR 719
 QY 65 IDQYNNSSSLRN-----HLNLEDDIMAQISITNGDK-----VEDYEKNIKKARKNL 111
 DB 720 MEQHRQHSLSQVQEAHLDRLLDLRQQS---DKETLAFHLEKVKDKLKNKRSRYECF 775
 QY 112 RVIASITKETWKIDSLDNLVKEV 134
 DB 776 HTL--LTKEVMEYPAI--MLKEL 794
 ID Q9S062 PRELIMINARY; PRT; 192 AA.
 AC Q9S062;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Conserved hypothetical protein.
 GN BBL35.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Plasmid cp32-8.
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 ON NCBI_TaxID=139;
 RN [1]


```

RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL: AJ414144; CAC89567.1; -.
DR InterPro: IPR000563; Flag_FliH.
DR Pfam: PF02108; FliH; 1.
DR PRINIS; PRO1003; FLGFLH.
KW Flagella; Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 26725 MW; 6088795DBE257B09 CRC64;

Query Match 11.8%; Score 91; DB 16; Length 239;
Best Local Similarity 26.5%; Pred. No. 13;
Matches 41; Conservative 25; Mismatches 57; Indels 32; Gaps 7;

QY 7 DNVNLLEBQYEL--GFKEGQIQGKDYLEGKEYGQTGFQRFLLIGYIOELMKFWLSH 64
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 35 DQTLDPAEYQKQLMAGFQEGISQGFQKGLAEGKEGEGYQGVV---LGHDDGLKK----- 85
QY 65 IDQYNNSSLRHNLNLEIDMAQIS--ITNGDKVEDEYKNIKARNKLRVIASITKETW 122
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 86 ----GRIEGRQSELASFNDVIKPSGYITQLHTYLETYEQ--RRRDELLQVKEKVTQVI 139
QY 123 KID-----SLDNLVKE-----VGGTLOVSENP 144
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 RCELALQPAQLTLVEEALAALEPMVPOQLKVLNLP 174

RESULT 15
Q8CKH8 PRELIMINARY; PRT; 247 AA.
AC Q8CKH8;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical.
GN Y3462.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Reese D.J., Mau B., Zhou S., Schwartz D.C.,
RA Retherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner P.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL: AF013949; AAM87011.1; -.
KW Hypothetical protein.
SQ SEQUENCE 247 AA; 27690 MW; 566022F0B496D008 CRC64;

Query Match 11.8%; Score 91; DB 16; Length 247;
Best Local Similarity 26.5%; Pred. No. 14;
Matches 41; Conservative 25; Mismatches 57; Indels 32; Gaps 7;

QY 7 DNVNLLEBQYEL--GFKEGQIQGKDYLEGKEYGQTGFQRFLLIGYIOELMKFWLSH 64
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 43 DQTLDPAEYQKQLMAGFQEGISQGFQKGLAEGKEGEGYQGVV---LGHDDGLKK----- 93
QY 65 IDQYNNSSLRHNLNLEIDMAQIS--ITNGDKVEDEYKNIKARNKLRVIASITKETW 122
DB : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 94 ----GRIEGRQSELASFNDVIKPSGYITQLHTYLETYEQ--RRRDELLQVKEKVTQVI 147
QY 123 KID-----SLDNLVKE-----VGGTLOVSENP 144
Db 148 RCELALQPAQLTLVEEALAALEPMVPOQLKVLNLP 182

Search completed: February 16, 2004, 14:14:12
Job time : 35 secs

```

Result No.	Score	Query			DB	ID	Description
		Match	Length				
1	223	28.9	198	1	YN00	YEAST	P53946 saccharomyc
2	97.5	12.6	610	1	MUTL	BORBU	O51229 borrelia bu
3	94	12.2	627	1	FLGK	BORBU	P70859 borrelia bu
4	91.5	11.9	1005	1	RA50	METUA	O58718 methanococc
5	87	11.3	1539	1	PY373	HUMAN	O15078 homo sapien
6	84.5	10.9	852	1	RA50	THEMA	O93131 thermotoga
7	83.5	10.8	879	1	RA50	PYRHO	O58687 pyrococcus
8	83.5	10.8	860	1	RA50	PYRAB	O58687 pyrococcus
9	83.5	10.8	930	1	TR66	MOUSE	G61143 mus musculus
10	82.5	10.7	1225	1	SMC1	YEAST	P32908 saccharomyc
11	82	10.6	1966	1	MY8B	CAEEL	P22566 caenorhabdi
12	81	10.5	320	1	K6P	SALRY	O8XG19 salmoneilla
13	81	10.5	521	1	KYP	BORBU	O51603 borrelia bu
14	81	10.5	564	1	M12	STRPV	P19401 streptococc
15	81	10.5	1453	1	PY373	BOVIN	O9TU23 bos taurus
16	81	10.5	2418	1	SPCA	HUMAN	P02549 homo sapien
17	80	10.4	1878	1	RA50	AQUAE	O67124 aquifex aeo
18	79.5	10.3	1699	1	DPOL	THEGS	O9H844 thermococcu
19	79	10.2	1173	1	ATC2	YEAST	P38929 saccharomyc
20	78.5	10.2	395	1	YU42	YEAST	P47122 saccharomyc
21	78.5	10.2	472	1	XISA	ANASP	P08862 anabaena sp
22	78.5	10.2	704	1	MEPB	RAT	P28826 rattus norv
23	78.5	10.2	714	1	YJL2	YEAST	P47025 saccharomyc
24	78.5	10.2	903	1	YB56	METUA	O58556 methanococc
25	78.5	10.2	956	1	SVI	AQUAE	O66651 aquifex aeo
26	78.5	10.2	2245	1	MYST	DICDI	P54697 dictyosteli
27	78.5	10.2	4486	1	DYH9	HUMAN	O9NYC9 homo sapien
28	78	10.1	382	1	SPCN	MOUSE	F16546 mus musculu
29	78	10.1	460	1	HENY	PSEAR	P77915 pseudomonas
30	78	10.1	638	1	YAB1	YEAST	P311376 saccharomyc
31	78	10.1	1421	1	YMP3	CAEEL	O10947 caenorhabdi
32	78	10.1	1939	1	CYAA	SACKL	P23466 saccharomyc
33	77.5	10.0	534	1	IF2P	SULAC	P95691 sulfolobus

RESULT 2

```

MUTL BORBU
ID MUTL BORBU STANDARD; PRT; 610 AA.
AC 051229;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR B02011.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
EX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,
RA Uterback T., Wathley L., McDonald L., Attiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586 (1997).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A001131; AAC66594.1; -
CC PIR; C70126; C70126.
CC HSP; P23367; LBKN.
CC TIGR; B02011; -
CC HAMAP; MF_00149; -; 1.
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR002099; DNA_mis_repair.
CC Pfam; PF01119; DNA_mis_repair; 1.
CC Pfam; PF02518; HATPase_C; 1.
CC SMART; SM00387; HATPase_C; 1.
CC TIGRFAMS; TIGR00585; mutL; 1.
CC PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
CC DNA repair; Complete Proteome.
KW SEQUENCE 610 AA; 71542 MW; 36A41A8A884EC9FA CRC64;
SQ
Query Match 12.6%; Score 97.5; DB 1; Length 610;
Best Local Similarity 24.7%; Pred. No. 1.4;
Matches 37; Conservative 27; Mismatches 51; Indels 35; Gaps 6;
QY 1 MSDDIDNVLNLE-----EVLGFKK-GQIGTKDQYLECKEYGYGTGFORPLIGYI 54
Dy 407 IGNIPSDNFLFEPPPKNKEKEIKENYIGI-----FSEFLIVEKI 448
QY 55 QELMKFWLSHIDQYNN-----SSLRNHLNLEDMAQISITNGKVED-YEKNIKAR 108
Dy 449 NEIY-----FIDQHAVHEKIIVEKLRSKNQKLVPIEFVVDKNIEEIDSEIEYK 503
QY 109 NKLRIASITKWTWKIDSLDNLVKEVGGTL 138
Dy 504 QMDIISKIGPKYQLESIPNICSQYENTL 533

```

```

RESULT 3
FLGK BORBU STANDARD; PRT; 627 AA.
AC P70859; O51199;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellar hook-associated protein 1 (HAP1).
GN FLGK OR B0181.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=212;
EX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman C.,
RA Uterback T., Wathley L., McDonald L., Attiach P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586 (1997).
CC -1- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U62901; AAB62740.1; -
CC PIR; A3001129; AAC66579.1; -
CC EMBL; AJ003222; CAA06002.1; -
CC PIR; E70122; E70122.
CC TIGR; B0181; -
CC InterPro; IPR001444; Flag_bb_rod.
CC InterPro; IPR002371; Flag_hookAp1.
CC PRINTS; PR01005; FLGHOOKAP1.
CC PROSITE; PS00588; FLAGELLA_BB_ROD; FALSE_NEG.
CC Flagella; Complete proteome.
KW CONFLICT 330 330 F -> S (IN REF. 1).
FT CONFLICT 551 551 A -> R (IN REF. 1).
FT CONFLICT 554 554 A -> R (IN REF. 1).
FT CONFLICT 558 558 A -> R (IN REF. 1).
FT CONFLICT 607 627 ASKFTIVSELIDTVINQGV -> QVNLSFLN (IN
FT REF. 1).
SQ SEQUENCE 627 AA; 70574 MW; 6D3B2F934900FF1 CRC64;
Query Match 12.2%; Score 94; DB 1; Length 627;
Best Local Similarity 25.7%; Pred. No. 2.7;
Matches 49; Conservative 23; Mismatches 51; Indels 68; Gaps 10;
QY 10 LMLEEQVELGFKK-----QIGTKQYV-----EGKEYGYGTGFORFLII----- 51
Dy 55 LNAKQKQGLG--QGVIVQSDIRVDELLNTRITIEESHLRGYWTSDKFTSILEVDVNEP 112
QY 52 ---GYIOELMKFWLSHIDQYNNSSL-----RNHLNLED--IMA--Q 87
Dy 113 EDCSIRKRLNDFWESWHDLANOQGLAEKILKRGKSCFGIIRFHSLEIYIMANDE 172

```


QY 88 ISITNGDKEVEDYERKNIKARNKLRVIAITKETWK-----IDSLDNLVKEVGG 136
 DB 173 IKITDE-----ANNYRIANLNKISQAMKNDPNLDNDARDLVVKEGLN 220
 QY 137 TLOVS-ENPDD 146
 DB 221 IISVSENKOD 231
 RESULT 4
 RASO_METUA STANDARD; PRT; 1005 AA.
 ID Q58718;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE DNA double-strand break repair rad50 ATPase.
 GN RAD50 OR MJ1322.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 RA Karavague A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.P., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 RL Science 273:1058-1073 (1996).
 CC -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
 CC rad50/mre11 complex possesses single-strand endonuclease activity
 CC and ATP-dependent double-strand-specific exonuclease activity.
 CC Rad50 provides an ATP-dependent control of mre11 by unwinding
 CC and/or repositioning DNA ends into the mre11 active site (By
 CC similarity).
 CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; U67572; AAB99331.1; -.
 CC FIR; AG4465; A64465.
 CC TIGR; MJ1322; -.
 CC
 CC HAMAP; MF_00449; -; 1.
 CC InterPro; IPR003593; AAA ATPase.
 CC InterPro; IPR003439; ABC transporter.
 CC InterPro; IPR003405; SMC_C.
 CC InterPro; IPR003395; SMC_N.
 CC Pfam; PF04423; Rad50_zn_hook; 1.
 CC Pfam; PF02483; SMC_C; 1.
 CC Pfam; PF02463; SMC_N; 1.
 CC Pfam; PF02463; SMC_N; 1.
 CC ProDom; PD000006; ABC transporter; 1.
 CC SMART; SM00382; AAA; 1.
 CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
 KW NP_BIND 32 39 ATP (By similarity).
 FT DOMAIN 158 849 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 11.9%; Score 91.5; DB 1; Length 1005;
 Best Local Similarity 24.7%; Pred. No. 7.3;
 Matches 43; Conservative 34; Mismatches 62; Indels 35; Gaps 7;
 QY 3 DID-IDNVLN-----LEEQYELGFKGQIQGTQOYLKGKYGQTGFQR 47
 DB 342 DIDNLDTLNKKIDBIERVETIKDLLEELKNLEIEIKYKICECKSY-----YEK 396
 QY 48 FL-IIGYIOELMKFWLSHDQYNNSSSLRNHLNLEIDIMAI-----SITNGDKE 96
 DB 397 YLELEKAVEYNKLTLEVTLLQEKKSIEKINDLETRINKLLEETKNIDIESIENSLKE 456
 QY 97 VEDYEK---NIKARNKL-RVIASTKETWKIDSLDNLVKEVGGTQVSENPDD 146
 DB 457 IEKKKVLNLOKEKIELNKLKLGELNLSIKRLKILDELKEVKGKCLCKTPID 510
 RESULT 5
 Y373_HUMAN STANDARD; PRT; 1539 AA.
 ID Q15078;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0373.
 GN KIAA0373.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=97349984; PubMed=9205841;
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. VII.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro.";
 RL DNA Res. 4:141-150 (1997).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC ENBL; AB002371; BAA20828.1; -.
 CC KW Hypothetical protein; Coiled coil.
 CC FT DOMAIN 18 1514 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;
 Query Match 11.3%; Score 87; DB 1; Length 1539;
 Best Local Similarity 20.8%; Pred. No. 27;
 Matches 31; Conservative 36; Mismatches 64; Indels 18; Gaps 4;
 QY 12 LEEQYELGFK-EGQ-----IQGTQOYLKGKYGQTGFQRFLI-----IGYI 54
 DB 302 LDEQALYARLEGRNRAKHLRQIQLSRQPSGALPLAQOEKFSKTMQLQNDKLKIM 361
 QY 55 GELMKFWLSHDQYNNSSSLRNHLNLEIDIMAIISITNGDKEVEDYKNIKAR-NKLRV 113
 DB 362 QEMKNSQOEHRNMENKTMELKLGLEELISTLTKDTNGAQKVINWNKLEELRLQELKL 421
 QY 114 IASITKETWKIDSLDNLVKEVGGTQVSE 142
 DB 422 NRELVKDEEIKYLNIISEYERTISLE 450
 RESULT 6
 RASO_THEME

RA50_THEME STANDARD; PRT; 852 AA.
Q9XIX1;
16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Probable DNA double-strand break repair rad50 ATPase.
RAD50 OR TW1636.
Thermotoga maritima.
Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
NCBI_TaxID=23336;
[1]
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heideberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329 (1999).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001806; AAD36703.1; -
DR PIR; D72230; D72230.
DR HAMAP; MF 00449; -; 1.
DR TIGR; TW1636; -
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
KW DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 155 711 COILED COIL (POTENTIAL).
SQ SEQUENCE 852 AA; 100001 MW; 31BA9F72A4EC5CD2 CRC64;
Query Match 10.9%; Score 84.5; DB 1; Length 852;
Best Local Similarity 21.2%; Pred. No. 22;
Matches 32; Conservative 35; Mismatches 49; Indels 35; Gaps 6;
QY 11 NLEEQYELGPKGQIQTKQYLEGKEYGYQTGFRLIGYQELMKFWLSHDQ-YN 69
DB 452 NIDFEF-KLDQKRSLENTLVNLRKK-----SLSSLLIEDLLMKIEGKK 496
QY 70 NSSSRNLNLEDDMAQISITNGKVEYDEYKNIKAKRLRV-----ASITKET 121
DB 497 NLKSRNQIKIEELHRLGYS-----EDLEKLDKRLKRLKIEERHSISQKITAD 550
QY 122 WKIDSLNLKVEVGTLQ-----VSENPDDM 147
DB 551 VOISQIENQLKEIGKEIEAKRETLKEOREM 591
RESULT 7
RA50_PYRHO STANDARD; PRT; 879 AA.
ID RA50_PYRHO
AC O58687;
DT 16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
Probable DNA double-strand break repair rad50 ATPase.
RAD50 OR TW1636.
Thermotoga maritima.
Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
NCBI_TaxID=23336;
[1]
SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RC MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heideberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329 (1999).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF001806; AAD36703.1; -
DR PIR; D72230; D72230.
DR HAMAP; MF 00449; -; 1.
DR TIGR; TW1636; -
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF02463; SMC_N; 1.
KW DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 155 711 COILED COIL (POTENTIAL).
SQ SEQUENCE 852 AA; 100001 MW; 31BA9F72A4EC5CD2 CRC64;
Query Match 10.9%; Score 84.5; DB 1; Length 852;
Best Local Similarity 21.2%; Pred. No. 22;
Matches 32; Conservative 35; Mismatches 49; Indels 35; Gaps 6;
QY 11 NLEEQYELGPKGQIQTKQYLEGKEYGYQTGFRLIGYQELMKFWLSHDQ-YN 69
DB 452 NIDFEF-KLDQKRSLENTLVNLRKK-----SLSSLLIEDLLMKIEGKK 496
QY 70 NSSSRNLNLEDDMAQISITNGKVEYDEYKNIKAKRLRV-----ASITKET 121
DB 497 NLKSRNQIKIEELHRLGYS-----EDLEKLDKRLKRLKIEERHSISQKITAD 550
QY 122 WKIDSLNLKVEVGTLQ-----VSENPDDM 147
DB 551 VOISQIENQLKEIGKEIEAKRETLKEOREM 591
RESULT 7
RA50_PYRHO STANDARD; PRT; 879 AA.
ID RA50_PYRHO
AC O58687;
DT 16-OCT-2001 (Rel. 40, Created)
16-OCT-2001 (Rel. 40, Last sequence update)
28-FEB-2003 (Rel. 41, Last annotation update)
DNA double-strand break repair rad50 ATPase.
RAD50 OR PH0329.
Pyrococcus horikoshii.
Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
Pyrococcus.
NCBI_TaxID=53953;
[1]
SEQUENCE FROM N.A.
RC STRAIN=OT3;
RC MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76 (1998).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000004; BAA30025.1; -
DR PIR; C71083; C71083.
DR HAMAP; MF 00449; -; 1.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR003405; SMC_C.
DR InterPro; IPR003395; SMC_N.
DR Pfam; PF04423; Rad50_zn_hook; 1.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR SMART; SM00382; AAA; 1.
KW DNA repair; Hydrolyase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 30 37 ATP (BY SIMILARITY).
FT DOMAIN 141 744 COILED COIL (POTENTIAL).
SQ SEQUENCE 879 AA; 103673 MW; FD4E30FD1B8CDB29 CRC64;
Query Match 10.8%; Score 83.5; DB 1; Length 879;
Best Local Similarity 19.8%; Pred. No. 27;
Matches 37; Conservative 40; Mismatches 61; Indels 49; Gaps 6;
QY 1 MSDIDIDNLEEQYELGFKS-----GOIQGTQDQYLEGKEYGYQT----- 43
DB 512 LSKINLED-LKROKEEYELLKSNLKGESVLSKEVSNELNDYKNESKLEIFDKAKK 570
QY 44 -----GFQRF-LIIGYQELMKFWLSHDQYANSSSLRNHNLEDDMAQI-- 88
DB 571 ELSIEDRLRLGFKTIDELSGRIRELEKPHNFKYIEAKNAEKELRDLFESLKDEREELDK 630
QY 89 -----SITNGDKVE-----DYKNIKAKRLRVASITKETWKIDSLNLVK 132
DB 631 AFEELAKIETDIEKVTSQLNELQKFDQKYBEKREKMKLSWEIKLETKLEELRRRD 690
QY 133 EVGGLQ 139
DB 691 EIKSTIE 697

```

RESULT 8
ID RA50 PYRAB STANDARD; PRT; 880 AA.
AC Q9U2C8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR PYRAB12200 OR PAR0812.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
EN [1]
SEQUENCE FROM N.A.
RP STRAIN=GES / Orsay;
RX PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ248286; CAB50131.1; -
CC PIR; F75103; F75103.
CC HAMAP; MF 00449; - 1.
CC InterPro; IPR003593; AAA ATPase.
CC InterPro; IPR003433; ABC transporter.
CC InterPro; IPR003395; SMC_N.
CC Pfam; PF04423; Rad50_2n_Hook; 1.
CC ProDom; PD000006; ABC transporter; 1.
CC SMART; SMC0382; AAA; 1.
CC DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
CC NP BIND 30 37 ATP (BY SIMILARITY).
CC FT DOMAIN 144 745 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 880 AA; 103970 MW; FDB177EC7E026479 CRC64;
Query Match 10.8%; Score 83.5; DB 1; Length 880;
Best Local Similarity 20.1%; Pred. No. 27;
Matches 38; Conservative 35; Mismatches 49; Indels 67; Gaps 8;
QY 11 NLREQYELGFKGQIGTKQY--LEGKEYGVOTGFQFP-----11----- 50
Db 517 NLE-----ELEQKRREFEGNENFKUGELGLDRLKIKALEGRKLIIEKVNKKEE 572
QY 51 -----IGY-----IQELMKFWLGHIDQYNNSSLRNHLNLEIDVAQIS-- 89
Db 573 LENLHRLQELGSPESVEELNLRIOLEEFHDKYVEAKKSESELRLKNLKEKETLQQA 632
QY 90 ---ITNGDKVEVDYENKIKARNKL-----RVIASITKETWKIDSLDLN 130
Db 633 FEMLADVENIEKKEAKLQLESKFNEEYEEKRELVLKLEFVSLTA---RLIELKYS 689

```

```

QY 131 VKEVGGTLQ 139
Db 690 VEQIKATLR 698
RESULT 9
ID TRP6 MOUSE STANDARD; PRT; 930 AA.
AC Q61143; Q922J1;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Short transient receptor potential channel 6 (TrpC6) (Calcium entry
channel).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
EN [1]
SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98037793; PubMed=9368034;
RA Boulay G., Zhu X., Peyton M., Jiang M., Hurst R., Stefani E.,
RA Birnbaumer L.;
RT "Cloning and expression of a novel mammalian homolog of Drosophila
transient receptor potential (Trp) involved in calcium entry secondary
to activation of receptors coupled by the Gq class of G protein.";
RL J. Biol. Chem. 272:29672-29680(1997).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=DBA/2;
RX MEDLINE=99158172; PubMed=10050885;
RA Buess M., Engler O., Hirsch H.H., Moroni C.;
RT "Search for oncogenic regulators in an autocrine tumor model using
differential display PCR: identification of novel candidate genes
including the calcium channel mtrp6.";
RL Oncogene 18:1487-1494(1999).
RN [3]
SEQUENCE OF 631-739 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=96234226; PubMed=8646775;
RA Zhu X., Jiang M., Peyton M., Boulay G., Hurst R., Stefani E.,
RA Birnbaumer L.;
RT "trp, a novel mammalian gene family essential for agonist-activated
capacitative Ca2+ entry.";
RL Cell 85:661-671(1996).
CC -!- FUNCTION: THOUGHT TO FORM A RECEPTOR-ACTIVATED NON-SELECTIVE
CC CALCIUM PERMEANT CATION CHANNEL. PROBABLY IS OPERATED BY A
CC PHOSPHATIDYLINOSITOL SECOND MESSENGER SYSTEM ACTIVATED BY RECEPTOR
CC TYROSINE KINASES OR G-PROTEIN COUPLED RECEPTORS. ACTIVATED BY
CC DIACYLGLYCEROL (DAG) IN A MEMBRANE-DELIMITED FASHION.
CC INDEPENDENTLY OF PROTEIN KINASE C. SEEMS NOT TO BE ACTIVATED BY
CC INTRACELLULAR CALCIUM STORE DEPLETION.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- TISSUE SPECIFICITY: LUNG AND BRAIN.
CC -!- PTM: N-GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO THE TRANSIENT RECEPTOR FAMILY. STRPC
CC SUBFAMILY.
CC -!- SIMILARITY: Contains 3 ANK repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U49069; AAC06146.1; -
CC EMBL; AF057748; AAC64394.1; -
CC MGD; MGI:109523; TrpC6.
CC GO; GO:0015279; F:store-operated calcium channel activity; IDA.

```

DR GO: 0007204; P: cytosolic calcium ion concentration elevation; IDA.
 DR InterPro: IPR002110; ANK.
 DR InterPro: IPR002111; Cat channel_TrpL.
 DR InterPro: IPR005821; Ion trans.
 DR InterPro: IPR002153; Trans receptor.
 DR InterPro: IPR004729; Trp channel.
 DR InterPro: IPR005462; Trp channel.
 DR Pfam: PF00023; ank; 4.
 DR Pfam: PF00520; ion trans; 1.
 DR PRINTS: PRO1097; TNSRRECEPTR.
 DR PRINTS: PRO1647; TRPCHANNEL6.
 DR SMART: SM00248; ANK; 2.
 DR TIGRFAMS: TIGR00870; trp; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 1.
 DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
 DR KW Ionic channel; Transmembrane; Ion transport; Calcium channel;
 FT ANK repeat; Repeat; Glycoprotein.
 FT DOMAIN 1 437 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 438 458 POTENTIAL.
 FT DOMAIN 459 486 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 487 507 POTENTIAL.
 FT DOMAIN 508 520 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 521 541 POTENTIAL.
 FT DOMAIN 542 591 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 592 612 POTENTIAL.
 FT DOMAIN 613 635 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 636 656 POTENTIAL.
 FT DOMAIN 657 705 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 706 726 POTENTIAL.
 FT DOMAIN 727 930 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 131 160 ANK 1.
 FT REPEAT 162 188 ANK 2.
 FT REPEAT 217 246 ANK 3.
 FT CARBOHYD 560 560 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 3 56 MISSING (IN REF. 2).
 FT CONFLICT 105 105 W -> A (IN REF. 2).
 FT CONFLICT 114 114 V -> R (IN REF. 2).
 FT CONFLICT 134 134 N -> D (IN REF. 2).
 FT CONFLICT 184 184 S -> A (IN REF. 2).
 FT CONFLICT 371 371 D -> Y (IN REF. 2).
 FT CONFLICT 436 437 PR -> RG (IN REF. 2).
 FT CONFLICT 905 905 S -> T (IN REF. 2).
 SQ SEQUENCE 930 AA; 106732 MW; CF2IA426972732F3 CRC64;
 Query Match 10.8%; Score 83.5; DB 1; Length 930;
 Best Local Similarity 25.3%; Pred. No. 29;
 Matches 37; Conservative 29; Mismatches 49; Indels 31; Gaps 6;
 Qy 19 LGKKEGQ---IQGTDQYLEGKEYGTGTFQRFLLITGYIQELMKP-----WLSHIDQYNN 70
 Db 777 LKFKKNCLELQGGKQFQDEAKNKNKNEKFKGSGSHEDLSKFLDKNQLAHNKQSST 836
 Qy 71 SSSLRHNLNLEDIMAGISITNGDKVEDYKNIKARNKLRVIASTTKETWKIDSLDNL 130
 Db 837 RSSSDYHLN-----SFSNPPRQ---YQIMKRLIKRYVLAQIDKESDEVN--EGE 882
 Qy 131 VKEVGGLT-----QVSENPPDM 147
 Db 883 LKEIKQDISLRVELLEKQNSGSDL 908
 RESULT 10
 SMCL_YEAST
 ID SMCL_YEAST
 AC P12908; STANDARD; PRT; 1225 AA.
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Structural maintenance of chromosome 1 (DA-box protein SMCL).
 GN SMCL OR CHL10 OR YFL008W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCB1_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., AND MUTANTS SMCL1-1 AND SMCL1-2.
 RX MEDLINE=94103320; PubMed=8276886;
 RA Strunnikov A.V., Lariouov V.L., Koshland D.;
 RT "SMCL: an essential yeast gene encoding a putative head-rod-tail
 protein is required for nuclear division and defines a new ubiquitous
 protein family";
 RL J. Cell Biol. 123:1635-1648 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C / AB972;
 RX MEDLINE=95400292; PubMed=7670463;
 RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
 RA Sasamura S.-I., Sasamura M., Tsuchiya Y., Soeda E., Yokoyama K.,
 RA Yamaraki M., Tashiro H., Eki T.;
 RT "Analysis of the nucleotide sequence of chromosome VI from
 Saccharomyces cerevisiae";
 RL Nat. Genet. 10:261-268 (1995).
 RN [3]
 RP IDENTIFICATION IN A COHESIN COMPLEX WITH SMC3; IRR1 AND MCD1, AND
 RP INTERACTION OF THE COHESIN COMPLEX WITH SCC2.
 RX PubMed=990856;
 RA Toth A., Ciosk R., Uhlmann F., Galova M., Schleiffer A., Nasmyth K.;
 RT "Yeast cohesin complex requires a conserved protein, Eco1p(Ctf7), to
 establish cohesion between sister chromatids during DNA replication.";
 RL Genes Dev. 13:320-333 (1999).
 RN [4]
 RP IDENTIFICATION IN A COHESIN COMPLEX WITH SMC3; MCD1 AND IRR1, AND
 RP STRUCTURE.
 RX PubMed=11983169;
 RA Haering C.H., Loewe J., Hochwagen A., Nasmyth K.;
 RT "Molecular architecture of SMC proteins and the yeast cohesin
 complex.";
 RL Mol. Cell 9:773-788 (2002).
 CC -!- FUNCTION: Involved in chromosome cohesion during cell cycle and in
 DNA repair. Central component of cohesin complex. The cohesin
 complex is required for the cohesion of sister chromatids after
 DNA replication. The cohesin complex apparently forms a large
 proteinaceous ring within which sister chromatids can be trapped.
 CC At anaphase, the complex is cleaved and dissociates from
 chromatins, allowing sister chromatids to segregate.
 CC -!- SUBUNIT: Cohesin complexes are composed of the SMCL1 and SMC3
 heterodimer attached via their hinge domain, MCD1/SCC1 which link
 them, and IRR1/SCC3, which interacts with MCD1. The cohesin
 complex also interacts with SCC2, which is required for its
 association with chromosomes.
 CC -!- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
 CC Before prophase it is scattered along chromosome arms. At
 anaphase, the MCD1 subunit of the cohesin complex is cleaved,
 CC leading to the dissociation of the complex from chromosomes,
 CC allowing chromosome separation.
 CC -!- DOMAIN: The flexible hinge domain, which separates the large
 intramolecular coiled coil regions, allows the heterotypic
 CC interaction with the corresponding domain of SMC3, forming a V-
 CC shaped heterodimer. The two heads of the heterodimer are then
 CC connected by different ends of the cleavable MCD1 protein, forming
 CC a ring structure.
 CC -!- SIMILARITY: Belongs to the SMC family. SMC1 subfamily.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L00602; AAA16595.1; --
 CC EMBL; D50617; BAA09230.1; --
 CC PIR; A49464; A49464.
 CC SGD; S0001886; SMCL1.
 CC GO; GO:0008278; C:cohesin complex; IDA.

DR GO; GO:0003680; F:AT DNA binding activity; IDA.
 DR GO; GO:0000217; F:DNA secondary structure binding activity; IDA.
 DR GO; GO:0003690; F:double-stranded DNA binding activity; IDA.
 DR GO; GO:0000070; P:mitotic chromosome segregation; IMP.
 DR InterPro; IPR003405; SMC C.
 DR InterPro; IPR003395; SMC N.
 DR Pfam; PF02463; SMC N; 1.
 DR Pfam; PF02483; SMC C; 1.
 KW Mitosis; Cell cycle; Chromosome partition; ATP-binding; Coiled coil;
 KW Nuclear protein.
 FT NP_BIND 33 40
 FT DOMAIN 173 489
 FT DOMAIN 490 678
 FT DOMAIN 679 1063
 FT DOMAIN 1057 1061
 FT DOMAIN 1129 1164
 FT MUTAGEN 173 173
 FT MUTAGEN 458 458
 SQ SEQUENCE 1225 AA; B504017AA0ECCAC8C CRC64;
 Query Match. 10.7%; Score 82.5; DB 1; Length 1225;
 Best Local Similarity 18.3%; Pred. No. 48;
 Matches 28; Conservative 36; Mismatches 50; Indels 39; Gaps 5;
 QY 5 DIDN---VLNIREQYELGFKGQIQGTQDQVLEKGYGTGQRFLLIYGIQELMKP 60
 DB 852 DLENAQVEMKSLSEQYAEIMKIGSIESKLEHK-----NHLDELQKK 894
 QY 61 WLSHIDQYNNSSLRNHLNLEDDIAQISITGDKK--VEDYE-----KNKKAR 108
 DB 895 FVTKQSELNSEDV-----LEDMSNLQVLRKRDGKEDLEKPLDRLVATLKNCKLSN 948
 QY 109 NKURVATISYKFTWKDLSLDNLVKEVGGTLOWS 141
 DB 949 INIPISSETTIDPLISSTDNEAITNSIDIN 991
 RESULT 11
 MYSB CAEL STANDARD; PRT; 1966 AA.
 AC P02566;
 DT 21-JUN-1986 (Rel. 01, Created)
 DT 21-JUN-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain B (MHC B).
 GN UNC-54 OR MYO-4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Ploterinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83273600; PubMed=6576334;
 RA Karn J., Brenner S., Barnett L.;
 RA "Protein structural domains in the Caenorhabditis elegans unc-54
 RT myosin heavy chain gene are not separated by introns."
 RL Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
 RN [2]
 RP SEQUENCE OF 850-1966 FROM N.A.
 RX MEDLINE=8272395; PubMed=7202124;
 RA McLachlan A.D., Karn J.;
 RA "Periodic charge distributions in the myosin rod amino acid sequence
 RT match cross-bridge spacings in muscle."
 RL Nature 299:226-231(1992).
 RN [3]
 RP SEQUENCE OF 1876-1966 FROM N.A.
 RX MEDLINE=8323892; PubMed=6571695;
 RA Wills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,
 RA Waterston R.H.;
 RA "The genes sup-7 X and sup-5 III of C. elegans suppress amber
 RT nonsense mutations via altered transfer RNA."
 RL Cell 33:575-583(1983).
 CC 1- FUNCTION: MUSCLE CONTRACTION.

CC 1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC 1- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC 1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HWM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC 1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
 CC C.ELEGANS.
 CC 1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
 CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
 CC 1- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; J01050; AAA28124.1; --
 CC EMBL; V01494; CAA24738.1; --
 CC HSP; P08799; LMWD.
 DR InterPro; IPR001609; myosin head.
 DR InterPro; IPR004009; Myosin N.
 DR InterPro; IPR002928; Myosin tail.
 DR Pfam; PF00063; myosin head; 1.
 DR Pfam; PF02736; Myosin N; 1.
 DR Pfam; PF01576; Myosin tail; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00242; MYSC; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family.
 FT DOMAIN 1 850
 FT DOMAIN 851 1164
 FT DOMAIN 1165 1176
 FT DOMAIN 1177 184
 FT NP_BIND 177 184
 FT DOMAIN 565 687
 FT DOMAIN 769 783
 FT MOD_RES 128 128
 FT MOD_RES 705 705
 FT MOD_RES 715 715
 FT CONFLICT 1337 1337 E -> R (IN REF. 2).
 FT CONFLICT 1880 1880 I -> L (IN REF. 2).
 SQ SEQUENCE 1966 AA; 225125 MW; B66F0B2FE27B67F CRC64;
 Query Match 10.6%; Score 82; DB 1; Length 1966;
 Best Local Similarity 20.9%; Pred. No. 89;
 Matches 37; Conservative 35; Mismatches 59; Indels 46; Gaps 6;
 QY 7 DNVNLLEBQVLEGFKEQIQGTQDQVLE-----GKEVGYGTGQRFLLIYGIQELM 58
 DB 968 DLEMSLRKAESEKQSKDQIRSLQDEMQQQDEATKLNKKKHEEINR-----KLM 1019
 QY 59 KFWLHIDQYNN-----SSSLRNHLNLEDDIM-----AQISIT 91
 DB 1020 EDLQSEEDKGNHKNVKAKLEQTLDLDDLSLREKRAADLDKQKRVGELKTAQENID 1079
 QY 92 NGDKVEDEYKNIKAKRNKLAVIAS-ITKWTWKDLSLDNLVKEVGGTLOVSENDDM 147
 DB 1080 ESGRQRHDLNLLKKSELSHVSRSLEDEQALYSKLQRQIKD--GQSRISELEEL 1134

```

DR  ProDom; P000707; Ppfuckinase; 1.
DR  PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
KW  Kinase; Transferase; Glycolysis; Complete proteome.
SQ  SEQUENCE 320 AA; 34915 MW; 68B0DDFCF689FA420 CRC64;

      Query Match      10.5%; Score 81; DB 1; Length 320;
      Best Local Similarity 25.2%; Pred. No. 14;
      Matches 35; Conservative 23; Mismatches 47; Indels 34; Gaps 7

QY  18  ELGKEGQIGTKDYLEGKEY--GYQTGFQRFRLIIGYIQELMKFWLSHIDQYNSSSLR 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db   115  EMGFPGTGLPGTIDNDIGDYITGY-----FTALGTVVEA-----IDRLRTSSSH 161

QY  76  NHLNLE-----DINMAQISITNGKVE-----EDYEKNKK--ARNKLEVIASI 117

Db  162  QRISIVFWGRCYCGDLFTLAAAIAGGCGEFTVWPEVEFNREDLVAELKAGIAGKKKEIAVAI 221

```

ID	SYK_BOBSU	STANDARD;	PRT;	521 AA.
AC	OS1603;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Lysyl-tRNA synthetase (EC 6.1.1.6) (Lysine-tRNA ligase) (LYSRS).			
DE	LYSRS OR B00459			
GN	Borrelia burgdorferi (Lyme disease spirochete).			
OS	Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.			
OC	Borrelia burgdorferi			
OX	NCBI_TaxID=139;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=ATCC 35210 / B31;			
RP	MEDLINE=9805943; PubMed=9403695;			
RY	Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,			
RA	Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,			
RA	Dougherty B., Tomb J.-F., Fleischmann R.J., Richardson D.,			
RA	Petersen J., Knelavage A.R., Quackenbush J.S., Salzberg S., Hanson M.,			
RA	Van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,			
RA	Uterback T., Wathey L., McDonald L., Artiach P., Bowman C.,			
RA	Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,			
RA	Smith H.O., Venter J.C.;			
RT	"Genomic sequence of a Lyme disease spirochaete, Borrelia			
RT	burgdorferi."			
RL	Nature 390:580-586(1997).			

```

[2] CHARACTERIZATION.
RP MEDLINE=98070760; PubMed=9405621;
RX Ibba M., Bono J.L., Rosa P.A., Soell D.;
RA "Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete
RT Borrelia burgdorferi."
RL Proc. Natl. Acad. Sci. U.S.A. 94:14383-14388(1997).
[3] CHARACTERIZATION.
RP MEDLINE=20570460; PubMed=11121028;
RX Soell D., Becker H.D., Plateau P., Blanquet S., Ibba M.;
RA "Context-dependent anticodon recognition by class I lysyl-tRNA
RT synthetases."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14224-14228(2000).
CC -!- CATALYTIC ACTIVITY: ATP + L-lysine + tRNA(Lys) = AMP + diphosphate
CC + L-lysyl-tRNA(Lys).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- MISCELLANEOUS: Able to charge E.Coli tRNA(Lys) in vitro.
CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AE001167; AAC67006.1; --
 DR PIR; B70182; B70182.
 DR TIGR; B80859; --
 DR HAMAP; MF_00177; --
 DR InterPro; IPR002904; Lys tRNA-synt_1c.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR Pfam; PF01921; tRNA-synt_1f; 1.
 DR TIGRfam; TIGR00467; lysS arch; 1.
 DR PROSITE; PS00178; AA tRNA_LIGASE I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 FT SITE 32 40 "HIGH" REGION.
 FT SITE 280 284 "KMSKS" REGION.
 SQ SEQUENCE 521 AA; 60938 MW; 6AF5A461AED3251D CRC64;

Query Match 10.5%; Score 81; DB 1; Length 521;
 Best Local Similarity 21.5%; Pred. No. 24;
 Matches 32; Conservative 22; Mismatches 49; Indels 46; Gaps 4;

QY 14 EEQVELGFKGQIGTQKDYLEKEGYQNGFORFLIIGYQELMKFWLSHIDQVNNSS 73
 DB 345 KEEKRAFK--RYVLSQPYNPKRIPYQVGFRLSVISCFE-----NNINK 390
 QY 74 LRNLNLEDDIMA-----QISITNGDKEVEDYENIKKARNKL 111
 DB 391 ILNLYKNVQDQDKLINKINCINWTRDFAPDFKLSRKFNDNWEILENSKKAINEL 450
 QY 112 -----RVIASITKETWKIDSLDNL 130
 DB 451 LDFLKNQFEVATEODIQNEIYKLSRENNI 479

RESULT 14
 ID M12_STRPY STANDARD; PRT; 564 AA.
 AC P19401;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE M protein, serotype 12 precursor (Fragment).
 GN EM12.
 OS Streptococcus pyogenes.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1314;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CS24 / Serotype M12;
 RX MEDLINE=88058777; PubMed=2445730;
 RA Robbins J.C., Spanier J.G., Jones S.J., Simpson W.J., Cleary P.P.;
 RT "Streptococcus pyogenes type 12 M protein gene regulation by upstream sequences."
 RL J. Bacteriol. 169:5633-5640(1987).
 CC -!- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES
 CC OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF
 CC THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO
 CC PHAGOCYTOSIS.
 CC -!- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -!- SIMILARITY: TO OTHER M PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M18269; AAA88573.1; --
 DR PIR; A40174; A60115.
 DR InterPro; IPR005877; Gpos YSIRK.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR InterPro; IPR006192; LPXTG.
 DR InterPro; IPR003345; M_repeat.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR Pfam; PF02370; M; 9.
 DR PRINTS; PR0015; GPOSANCHOR.
 DR TIGRfam; TIGR01168; YSIRK signal; 1.
 DR PROSITE; PS00847; GRAM_POS_ANCHORING; 1.
 KW Virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
 KM Antigen; Coiled coil; Signal.
 FT SIGNAL 1 41
 FT CHAIN 42 545 M PROTEIN, SEROTYPE 12.
 FT PROPEP 546 564 REMOVED BY SORTASE (POTENTIAL).
 FT DOMAIN 44 505 COILED COIL (POTENTIAL).
 FT SITE 542 546 GLY/PRO-RICH (CELL WALL-SPANNING).
 FT MOD_RES 545 545 LPXTG SORTING SIGNAL (POTENTIAL).
 FT NON_TER 564 564 AMIDS-LINKED TO CELL WALL (POTENTIAL).
 SQ SEQUENCE 564 AA; 62904 MW; 5F1549DACAA77B46 CRC64;

Query Match 10.5%; Score 81; DB 1; Length 564;
 Best Local Similarity 22.6%; Pred. No. 26;
 Matches 35; Conservative 30; Mismatches 48; Indels 42; Gaps 7;

QY 13 EEEQVE-LGFKEGQIGTKD---QYLEKEGYQNGFORFLIIGYQELMKFWLSHIDQ 67
 DB 49 EQKRLDQKFERLKQSELYLQYYDKNKSGYKGDW-----YVQQL-KMLNRDLEQ 100
 QY 68 YNNSSSLNHLNLEDDIMAQISITNGD-----KEVEDYENIK 104
 DB 101 AYNELSGRAH-----KDALGKLIDNADLKAKITELKSEVKEKNDVLSQIKKELEAEKDI 156
 QY 105 KKARNKLRIASITKETWKIDSLDNLKVEGGTTLQ 139
 DB 157 QFGREV--HAADLLRHKQEIAEKENVSKINGELQ 189

RESULT 15
 ID Y373_BOVIN STANDARD; PRT; 1453 AA.
 AC Q9TU23;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA0373.
 GN KIAA0373.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Trachea;
 RA Jovov B., Ripoll P.J., Benos D.J.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AF176816; AAF00990.1; --
 KW Hypothetical protein; Coiled coil.
 FT DOMAIN 37 1426 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF4139 CRC64;

Query Match 10.5%; Score 81; DB 1; Length 1453;
Best Local Similarity 20.1%; Pred.No. 76;
Matches 30; Conservative 34; Mismatches 67; Indels 10; Gaps 4;
QY 12 LEEBOYELGFK--EGQ-----IOGTQDYLEGKEYGYQTGFORFLI-----IGYI 54
DB 321 LDEKQALFYARLEGRNEAKHLRQTIOSLRRQFSGALPLAQOEKFSKTMQLQNDKIKM 380
QY 55 QELMKFWLSHIDQYNNSSSLRNHLNLEIDMAQISITNGDKEVEDYKNIKAR-NKLRY 113
DB 381 EEMKNSQOEHRSLKNKTLMEBELKLGLEDLSTIKDARGAQKVISMHTKIEELRLOELKL 440
QY 114 TASITKETWKIDSLDNLVKEVGGTLQVSE 142
DB 441 NRELVKDEEIKYLNLIIEVENTISSLE 469

Search completed: February 16, 2004, 14:12:32
Job time : 13 secs